AD-A104 043 MASSACHUSETTS INST OF TECH CAMBRIDGE ARTIFICIAL INTE--ETC F/G 12/1
THE IMAGE IRRADIANCE EQUATION: ITS SOLUTION AND APPLICATION.(U)
JUN 81 A R BRUSS N00014-80-C-0505
UNCLASSIFIED AI-TR-623 NL

1 OF 2
AD
A104 043

AI-TR-623

AD A104043

THE IMAGE IRRADIANCE EQUATION: ITS SOLUTION AND APPLICATION

ANNA R. BRUSS

JUNE 1981

D

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

ARTIFICIAL INTELLIGENCE LABORATORY

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER AI-TR-623	2. GOVT ACCESSION NO. AD-A104043
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) The Image Irradiance Equation: Its Solution and Application	5. TYPE OF REPORT & PERIOD COVERED (9) Technical Report,
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Anna R. Bruss	8. CONTRACT OR GRANT NUMBER(s) N00014-80-C-0505, N00014-77-C-0389
9. PERFORMING ORGANIZATION NAME AND ADDRESS Artificial Intelligence Laboratory 545 Technology Square Cambridge, Massachusetts 02139	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Blvd Arlington, Virginia 22209	12. REPORT DATE (11) Jun[redacted]981
	13. NUMBER OF PAGES 109
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Office of Naval Research Information Systems Arlington, Virginia 22217	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Image Irradiance Equation
Shape from Shading
Computer Vision

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

How much information about the shape of an object can be inferred from its image? In particular, can the shape of an object be reconstructed by measuring the light it reflects from points on its surface? These questions were raised by Horn [HO70] who formulated a set of conditions such that the image formation can be described in terms of a first order partial differential equation, the image irradiance equation. In general, an image irradiance equation has infinitely many solutions. Thus constraints necessary to find

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. a unique solution need to be identified. First we study the continuous image irradiance equation. It is demonstrated when and how the knowledge of the position of edges on a surface can be used to reconstruct the surface. Furthermore we show how much about the shape of a surface can be deduced from so called singular points. At these points the surface orientation is uniquely determined by the measured brightness.

Then we investigate images in which certain types of silhouettes, which we call b-silhouettes, can be detected. In particular we answer the following question in the affirmative: Is there a set of constraints which assure that if an image irradiance equeation has a solution, it is unique? To this end we postulate three constraints upon the image irradiance equation and prove that they are sufficient to uniquely reconstruct the surface from its image. Furthermore it is shown that any two of these constraints are insufficient to assure a unique solution to an image irradiance equation. Examples are given which illustrate the different issues.

Finally, an overview of known numerical methods for computing solutions to an image irradiance equation are presented.

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under contract number N00014-80-C-0505 and in part by the Office of Naval Research under contract number N00014-77-C-0389.

THE IMAGE IRRADIANCE EQUATION: ITS SOLUTION AND APPLICATION

by

Anna R. Bruss

Massachusetts Institute of Technology

June 1981

Revised version of a dissertation submitted to the Department of Electrical Engineering and Computer Science on December 31, 1980 in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

Abstract

How much information about the shape of an object can be inferred from its image? In particular, can the shape of an object be reconstructed by measuring the light it reflects from points on its surface? These questions were raised by Horn [HO70] who formulated a set of conditions such that the image formation can be described in terms of a first order partial differential equation, the *image irradiance equation.* In general, an image irradiance equation has infinitely many solutions. Thus constraints necessary to find a *unique* solution need to be identified.

First we study the continuous image irradiance equation. It is demonstrated when and how the knowledge of the position of edges on a surface can be used to reconstruct the surface. Furthermore we show how much about the shape of a surface can be deduced from so called *singular points.* At these points the surface orientation is uniquely determined by the measured brightness.

Then we investigate images in which certain types of silhouettes, which we call *b-silhouettes*, can be detected. In particular we answer the following question in the affirmative: Is there a set of constraints which *assure that if* an *image* irradiance equation has a solution, it is unique? To this end we postulate three constraints upon the image irradiance equation and prove that they are sufficient to uniquely reconstruct the surface from its image. Furthermore it is shown that any two of these constraints are insufficient to assure a unique solution to an image irradiance equation. Examples are given which illustrate the different issues.

Finally, an overview of known numerical methods for computing solutions to an image irradiance equation are presented.

Thesis Supervisor: Berthold K. P. Horn
Title: Associate Professor of Computer Science and Electrical Engineering

Acknowledgements

I would like to thank my thesis advisor Professor Berthold K. P. Horn for all the time he spent with me discussing this topic and for the encouragement he gave me during my years as a graduate student. I would also like to thank Professor Richard Melrose for contributing his ideas so generously. However, most of all I want to thank him for giving me the confidence to keep on working.

The comments of my readers Dr. Shimon Ullman and Professor Patrick H. Winston and of Peter Blicher, Dr. Mike Brady, Dr. Jeanne Ferrante and Stephen M. Trilling improved the style of this thesis. I really appreciate that they carefully read through drafts of this thesis. I would also like to thank Professor Hal Abelson for the valuable suggestions he made.

This thesis would never have been written without the stimulating environment of "Tech Square". I want to thank all of my colleagues throughout the years for the numerous hours they spent talking and listening, in particular Peter Andreae, Ken Forbus, Tomas Lozano-Perez, Barry Trager and Richard Zippel. I also want to thank Richard Zippel for his help in formatting this thesis and Bill Silver for the creation of the plotting routines.

The steady moral and financial support of my father Jan Bruss enabled me to enjoy my days as a graduate student and I want to thank him for everything by dedicating this thesis to him.

I want to thank my sister Elisabeth Saunders for making me apply to M.I.T. in the first place. I could not have made it without her and her husband Ronald Saunders' love and understanding.

All the ups and downs of the last five years I shared with a very special friend, Carol Schrager. Thanks for always being there!

I also want to thank Len Adleman and Jeanne Ferrante for their wonderful friendship which I always relied on and Jerry Barber, Barry Trager, Stephen Trilling, Lucia Vaina and Richard Zippel for helping me to keep my sanity throughout.

Last but not least, I want to thank David Shaw for sharing the pain and the happiness.

Contents

Chapter I

Motivation

How much information about the shape of an object can be inferred from its image? We are interested in a special aspect of this question: the *reconstruction problem*, which is to determine the shape of an object from measurements of the light reflected from its surface. In general, there are many surfaces which can give rise to the same image. So, we will try to identify and analyze constraints such that the shape of a surface can be *uniquely* reconstructed from its image. Our search for these constraints is guided by the following question: Are there any "properties" of an object that are also "properties" of its image or vice versa? In fact, as we shall see, such properties do exist.

Our work is based on Horn's thesis [HO70]. He formulated a set of conditions (to be discussed in the next chapter) which lead to a relation between the perceived brightness of a small patch of a surface and its normal vector. This relation, the *image irradiance equation*, is a first order partial differential equation (abbreviated in the following by FOPDE) and each of its solutions determines the shape of an object. The problem of finding solutions to the image irradiance equation is referred to in the literature as the *shape from shading problem*.

We will take two approaches towards finding a solution to the shape from shading problem termed as the *local* and the *global* approach. By the local approach we mean that only a small patch of an image is used to determine the shape of a surface. To the contrary, in the global approach we examine images in which a silhouette can be detected (here we refer to the outline of an image as a silhouette).

Intuitively, it seems clear that by looking at an image in which a silhouette can be identified we should be able to conclude more about the shape of a surface whose image we are analyzing than by just looking at a little patch. We will show that from certain

images which contain a silhouette we can uniquely infer the shape of the surface which gives rise to that image. Unfortunately the global approach is not always satisfactory; there are also many images containing silhouettes which could be the images of infinitely many different surfaces. There are also infinitely many surfaces which locally look the same. So we will determine conditions under which the global approach is better than the local approach. Notwithstanding, one can *sometimes draw interesting conclusions* about the shape of surfaces which give rise to the same image by just looking at a small patch of this image.

The local approach is taken to an extreme when we pose the following question: What can be deduced about the shape of a surface from so-called *singular points* of an image irradiance equation? At these points the surface normal to all solutions to such an equation is uniquely determined by the brightness there. We investigate the above stated question for a certain class of image irradiance equations, the so-called eikonal equations, which describe a variety of physical phenomena. For instance, experimental data suggest that the flux of secondary electrons in a scanning electron microscope can be described by an eikonal equation [LAWH60]. By using these secondary electrons to modulate the appropriate devices, an image of a surface is created by the microscope. Such an image exhibits shading [HO70, pp.85–87] and therefore to determine the shape of a surface from its image one effectively has to solve an eikonal equation. Studying eikonal equations, we show that the absolute value of the Gaussian curvature at a singular point of all surfaces which give rise to a particular image, is the same. Furthermore, assuming that the surface is convex at a singular point, we show that its shape can be uniquely determined in some neighborhood of such a point from the image intensities alone.

The other aspect of the shape from shading problem which we explore is its solution when the image contains a *b-silhouette* (which is defined below). In this case a global approach is taken. Let us first define the *bounding contour* of a surface: a point P is on the bounding contour if the line connecting the viewer and P *grazes* the surface (i.e., if this line lies in the tangent plane of P). Furthermore we assume that no two parts of a surface obscure each other, i.e., we assume that the bounding contour is not an occluding contour. The image (assuming orthographic projection) of a bounding contour will be called the *b-silhouette*. The surface normal at a point on a bounding contour is parallel to the normal vector to the b-silhouette and both vectors lie in the same plane. Thus, some or all of the first order partial derivatives of the function defining the surface are infinite for points on the bounding contour (we will say that some components of the surface gradient are *singular* along a curve).

Remark: In the context of this report the phrase *a function $f(x,y)$ is singular at a point (x_0, y_0)* will always mean that:

$$\lim_{\substack{x \to x_0 \\ y \to y_0}} f(x,y) = \pm\infty. \tag{I.1.1}$$

This terminology should not be confused with the notion of a *singular solution* which we will explain in section A.4.

For example, the bounding contour of a hemisphere lying on a plane B is a circle. Consider a Lambertian surface which has the property that each surface patch appears equally bright from all viewing directions. If we look at a Lambertian hemisphere such that the viewer and the light source are at the same point, its b-silhouette can be determined from its image. In this case the image irradiance equation describing the imaging situation is singular and *all* the surfaces which satisfy such an equation have a bounding contour. (We extend the notion of a singular function to equations.)

On the other hand, the existence and the position of a b-silhouette cannot be determined from a continuous image irradiance equation. Thus the existence of a surface which satisfies a continuous image irradiance equation and which has a bounding contour does not necessarily imply that all surfaces which satisfy this equation have a bounding contour.

In this report, Horn's work on the shape from shading problem is extended. He studied primarily images of smooth surfaces where the imaging situation can be described by a *continuous* image irradiance equation (which is defined rigorously in section III.1). Yet some objects have edges and their surface normals assume different values depending upon which side an edge is approached from (we will say that the surfaces normals are *discontinuous* along an edge). Can an edge be invisible in an image? In other words, is it possible that a surface has an edge without it producing a discontinuity in the equation? As it turns out, discontinuous solutions can arise from continuous equations. In this case *initial conditions* (discussed in sections III.2, A.6 and A.7) provide information about the occurrence of a discontinuity. Additionally, we will show (in section III.2.3) under which conditions edges on a surface can be used as such initial data.

We keep in mind that the image irradiance equation describes a physical situation and therefore the only solutions considered are those corresponding to (piecewise) smooth surfaces. (A rigorous definition of the notion *smooth surface* is given in section III.1.) However there are image irradiance equations for which no solution corresponding to a smooth surface exists. In particular, the existence of such a solution to a singular image irradiance equation is not guaranteed. We are primarily concerned with the identification of constraints which allow one to solve the reconstruction problem *uniquely*.

In general, a PDE describes a class of processes rather than a particular one. Consider, for example, the Laplace equation:

$$\Delta f = 0 \tag{I.1.2}$$

where Δ denotes the Laplace operator and f a scalar field. This PDE constrains the sources and the curl of the field f to be zero, but an infinite number of different fields exist which fall into this category. Only when some further conditions about f are specified can the solution to the Laplace equation be restricted to a single one.

Similarly, there are an infinite number of different surfaces which satisfy a given image irradiance equation. Thus, as Horn [HO70, HO75] has already observed in

general the image irradiance equation alone is not sufficient to solve the reconstruction problem uniquely. It remained an open question whether there are *any* imaging situations for which every surface gives rise to a different image. As indicated before, taking the global approach towards finding a solution to the shape from shading problem allows us to answer this question affirmatively. To this end we will postulate three constraints upon the image irradiance equation and prove that if these constraints are known to hold, the reconstruction problem can be solved uniquely.

We now briefly describe the different chapters in this thesis.

Chapter II is a summary of the issues involved in the shape from shading problem. For a more extensive discussion of this material see [HO70], [HO75] and [WOOD78].

Chapter III gives an overview of the different mathematical problems involved in solving a FOPDE. The mathematical details used in this chapter can be found in appendix I. In section III.1 we define two classes of image irradiance equations, the continuous and the singular equations. Section III.2 deals in detail with the continuous image irradiance equation. In particular we exhibit in section III.2.2 that certain continuous image irradiance equations can be transformed into singular equations. This implies that all surfaces which satisfy such an image irradiance equations have a bounding contour. In section III.2.3 we examine how edges on a surface can be used to reconstruct the surface. To help us understand the variety of integral surfaces of an image irradiance equation we introduce in section III.2.4 some concepts from differential geometry and in section III.2.5 gradient space as popularized by Mackworth [MAC73] and Horn [HO77]. Section III.3 describes the basic issues involved in solving a singular image irradiance equation. In section III.3.1 Marr's [MA77] work on occluding contours is reviewed. Section III.3.2 discusses the method of characteristic curves for singular image irradiance equations.

In chapter IV we show how a singular point of an eikonal equation constrains its possible solutions.

In chapter V we postulate three constraints upon an image irradiance equation such that the reconstruction problem can be solved uniquely. In section V.1 it is proven that if an image irradiance equation satisfies these constraints it has a unique solution. It is shown in section V.2 that any two of these constraints are insufficient to assure a unique solution to an image irradiance equation.

Chapter VI gives a brief description and discussion of known numerical methods for computing solutions to an image irradiance equation.

Chapter VII summarizes the results of this report and suggests some possible applications.

Chapter II

The Shape from Shading Problem

There are basically three components to the shape from shading problem which have to be taken into account. They are the light source, the object and the camera as depicted in figure 1 which is taken from [WOOD78, p.32], and termed as an *imaging configuration*. Henceforth we will assume that an image of a surface is produced by a camera. The shading of such an image can be explained as follows: The exposure of film in a camera (for fixed shutter speed) is proportional to *image irradiance*, the light flux per unit area falling on the image plane. Similarly, grey levels measured in an electronic imaging device are quantized measurements of image irradiance. It can be shown that image irradiance in turn is proportional to *scene radiance*, the light flux emitted by the object per unit projected surface area per unit solid angle [HOS79]. The factor of proportionality depends on details of the optical system, including the effective f-number.

Scene radiance depends on the

- surface material and its microstructure,
- the incident light flux, and,
- the orientation of the surface.

Now we want to relate the shape of a surface to the shading of its image. Consider a viewer-oriented coordinate system with the viewer located far above the surface on the z-axis. If the objects imaged are small compared to their distance from the viewer, one can approximate the imaging situation by an orthographic projection:

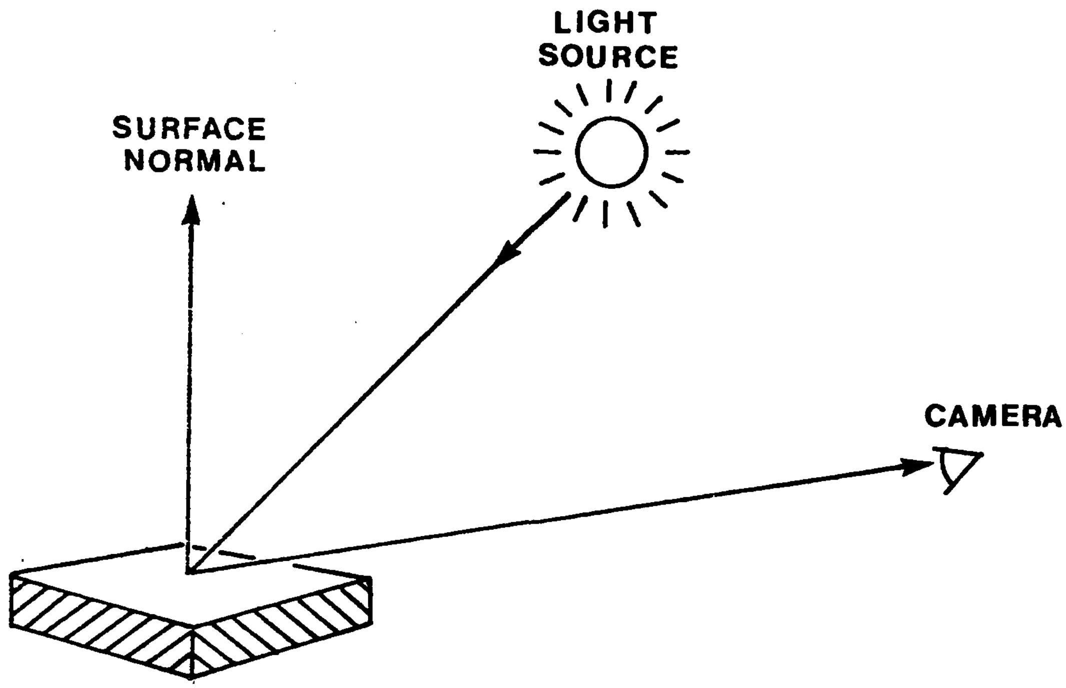


Figure 1. Imaging *configuration*

$$\tilde{x} = \frac{xf}{z_0} \qquad \tilde{y} = \frac{yf}{z_0} \tag{II.1.1}$$

where $(\tilde{x}, \tilde{y})$ are the coordinates of the image of a point (x, y, z) made with a system of effective focal length f, and the viewer is at a distance z_0 above the origin. We assume that $(x^2 + y^2 + z^2) \ll z_0^2$. For simplicity and without loss of generality it is also assumed that the viewing direction coincides with the z-axis.

The orientation of a patch of a surface can be specified by its gradient $(p, q, -1)$, where p and q are the first order partial derivatives of z with respect to x and y. For a given surface material and known incident light flux, scene radiance will depend only on surface gradient. The function which describes this dependence, $R(p, q)$, (or a contour representation in gradient space), is called the reflectance map.

Recall that image irradiance and scene radiance are proportional and that we assume orthographic projection. If $E(x, y)$ is the observed image irradiance at the point $(\tilde{x}, \tilde{y})$ in the image then:

$$R(p, q) = E(x, y) \tag{II.1.2}$$

where (p, q) are two components of the gradient at the corresponding point on the object

being imaged. This equation is called the *image irradiance equation*. It is clearly a first order partial differential equation since it involves only the first order partial derivatives p and q and the coordinates x and y. In summary, to derive an image irradiance equation we have to know the reflectivity function and the geometry relating the light source, the object and the camera.

How can an image irradiance equation be used to analyze images? Informally, suppose enough information about the imaging situation is known so that the reflectance map can be derived. Then at every point (x_0, y_0) the image irradiance denoted by E_0 can be measured. Using the image irradiance equation we can determine the possible values for p and q such that:

$$R(p,q) = E_0. \tag{II.1.3}$$

However, the values for p and q cannot be chosen independently as p and q have to be the first order partial derivatives of a function $z = z(x, y)$ defining a smooth surface. In general there are many values for p and q which satisfy equation (II.1.3) and represent the components of a surface gradient. Consequently, an image irradiance equation has many solutions or equivalently, for a fixed imaging configuration many surfaces will give rise to the same image. Thus we shall determine constraints under which the solutions to an image irradiance equation are restricted to a unique one.

A word of caution: Several issues such as mutual illumination, shadows or specularity are not addressed here.

Chapter III

Overview

III.1. Classification of the Image Irradiance Equation

In this chapter, we will further investigate the shape from shading problem. Our interests are twofold:

1) We wish to find the solutions to an image irradiance equation.

2) We wish to determine constraints which guarantee a unique solution to an image irradiance equation.

The results concerning these two issues are different for continuous and singular image irradiance equations, both of which are defined later in this section. First we make some general observations concerning an such equations and their solutions.

An image irradiance equation is a first order partial differential equation in two variables x and y. We are looking for a solution $z = z(x, y)$ (also called *integral surface*) which is a function of x and y and defines a surface from which light is reflected. To be precise, the class of functions which we allow as solutions to the FOPDE has to be specified and we proceed now with some relevant definitions. Using standard nomenclature, a function $f(x, y)$ is said to be of class C^k if it has continuous k-th order partial derivatives. The following definition captures formally our geometrical intuition about a surface in $\Re^3$:

Definition: Let $B \subseteq \Re^2$ be a connected and closed region, and let $z{:}B \mapsto \Re^3$ be continuous on B. Then the single-valued function $z = z(x, y)$ defines a *smooth surface* if z is C^1 at every interior point of B.

For smooth surfaces the unit *surface normal* is well defined for every point on the surface by:

$$\frac{(1, 0, \frac{\partial z}{\partial x}) \times (0, 1, \frac{\partial z}{\partial y}).}{\| (1, 0, \frac{\partial z}{\partial x}) \times (0, 1, \frac{\partial z}{\partial y}) \|} \tag{III.1.1}$$

A *piecewise smooth* surface is defined similarly. Unless otherwise stated we will always assume that the solutions to any given FOPDE are (piecewise) smooth surfaces. We will denote by p and q the first order partial derivatives of $z = z(x, y)$ with respect to x and y.

We shall exploit the following two facts about an image irradiance equation:

1) The equation does not depend explicitly on z.

2) The equation involves only two functions R and E, such that R depends only on p and q and E depends only on x and y.

An immediate consequence of 1 is that there are no singular solutions to the equation, i.e., the complete integral describes all possible solutions. (For an explanation of these terms see section A.4). As for 2, we will study image irradiance equations which fall into either one of the following two categories:

1) The functions R and E are C^1. We say that such an image irradiance equation is *continuous*. This case is discussed in section III.2.

2) The function R is C^1. The function E is singular in x and/or y but for all points (x, y) at which $E(x, y)$ assumes finite value, it is C^1. We say that such an *image irradiance equation* is *singular*. This case is discussed in section III.3.

Since the *measured* image irradiance is always finite, it seems at first that singular image irradiance equations are not of practical interest. However, for our purposes we can think about such equations as useful mathematical constructs, i.e., a singular image irradiance equation can be obtained by transforming a continuous image irradiance equation appropriately. By *appropriately* we mean that the transformation is one-to-one and onto and that the solutions to the original equation and the transformed one are the same (section III.2.2). For example, there is a transformation between the

continuous image irradiance equation (III.1.2) and the singular equation (III.1.3):

$$\frac{p^2+q^2}{1+p^2+q^2} = x^2+y^2 \qquad \text{(III.1.2)}$$

$$p^2+q^2 = \frac{x^2+y^2}{1-(x^2+y^2)}. \qquad \text{(III.1.3)}$$

We still have to explain why such a transformation is useful, i.e., why it makes more sense for us to analyze the singular equation (III.1.3) instead of the continuous equation (III.1.2). The reason is that we can gain some information about the integral surfaces of a singular image irradiance equation without first solving it. To this end let us define the terms bounding contour and b-silhouette:

Definition: A point P lies on the *bounding contour* of a surface defined by $z = z(x, y)$ if the tangent plane at P is perpendicular to the x-y plane. The *b-silhouette* is the orthographic projection of the bounding contour onto the x-y plane.

We will show below that *each* integral surface of any singular image irradiance equation has a bounding contour, a fact which is not true for continuous image irradiance equations. In section III.2.2 we will construct two integral surfaces of a continuous image irradiance equation such that one of them has a bounding contour whereas the other does not. Why do all integral surfaces of a singular image irradiance equation have a bounding contour? Recall (chapter I) that for points on the bounding contour p and/or q become infinite. Now, an image irradiance equation is singular if there are points (x_0, y_0) such that the *only* values for p and q which satisfy the equation at (x_0, y_0) are infinite. This explains the previously stated question. Furthermore, as the b-silhouette is the projection of the bounding contour onto the image plane, the points which constitute the b-silhouette can be uniquely determined from a singular image irradiance equation (section III.3).

We will not be concerned with discontinuous reflectance maps since these are rare. They may occur when dealing with specularities, an issue not addressed in this report. (It is also assumed that the reflectance map is not a constant since then the image will also be constant.)

III.2. The Continuous Image Irradiance Equation

In the mathematical literature, the most studied FOPDE's are continuous and have continuous first order partial derivatives. An overview of known results in this area can be found in appendix I and we summarize them in the next few sections only very briefly. In particular we explain why, in general, additional information is needed to restrict the solutions to a given image irradiance equation to a single one. Furthermore we address the problem of whether one can deduce any properties of the

integral surfaces of a given image irradiance equation by just inspecting the equation. To this end we investigate singular points, at which the tangent plane to all integral surfaces of an image irradiance equation is uniquely defined by the image intensity (section III.2.2). We can then show (chapter IV) that for a class of image irradiance equations the absolute value of the curvature of the integral surfaces at a singular point is uniquely defined by the measured brightness there. In addition, we discuss a class of image irradiance equations for each member of which we can deduce that all of its integral surfaces have a bounding contour.

The main tool for solving a FOPDE is the construction of so called *characteristic curves*. The following is a short explanation of this notion. A more detailed exposition can be found in sections A.2 and A.3. Let

$$R(p,q) = E(x,y) \tag{III.2.1}$$

be a continuous image irradiance equation. What kind of information about an integral surface (defined by $z = z(x,y)$) of the previous equation can be deduced from it? For a given point P on z the quantities p and q are constrained by (III.2.1) to lie on a curve. Since p and q determine the normal $(p,q,-1)$ at P, equation (III.2.1) constrains the feasible tangent planes at P to a one-parameter family, "enveloping a conical surface with P as vertex, called the *Monge cone*" [COHI62b, p.75]. The directions of the *generators* of a Monge cone are called *characteristic directions*. Now, the *characteristic curves* are those curves on an integral surface which at every point have as their tangent direction a characteristic direction. The characteristic curves can be determined by solving the characteristic equations which are five ordinary differential equations whose solutions depend on initial values:

$$\begin{aligned} \frac{dx}{ds} &= F_p \qquad \frac{dy}{ds} = F_q \qquad & \frac{dz}{ds} &= pF_p + qF_q \\ \frac{dp}{ds} &= -(pF_z + F_x) & \frac{dq}{ds} &= -(qF_z + F_y). \end{aligned} \tag{III.2.2}$$

Equivalently, we can say that for distinct initial values, different characteristic curves are determined. Since each integral surface of a FOPDE is swept out by characteristic curves (which we prove in section A.3), initial conditions (which are discussed further in the next sections) are necessary to restrict the solutions to an image irradiance equation to a single one; in the case where an image irradiance equation is continuous, a unique solution to the problem can be found provided that an appropriate initial strip defined by $x = x(t), y = y(t), z = z(t), p = p(t)$ and $q = q(t)$ is known (section A.6 and A.7). By *appropriate* we mean that this strip is not a characteristic strip and that the following condition holds:

$$\Delta \equiv \frac{dx}{ds}\frac{dy}{dt} - \frac{dy}{ds}\frac{dx}{dt} \neq 0. \tag{III.2.3}$$

Depending upon the initial curve, the integral surface is either continuous or discontinuous as will be discussed in greater detail later on.

III.2.1. The Linear Continuous Image Irradiance Equation

We first review the case where the image irradiance equation is a linear FOPDE e.g.,:

$$ap + bq = E(x, y) \tag{III.2.4}$$

where a and b are constants. This equation is of practical interest as its linear reflectance map describes the reflectivity properties of the maria of the moon where the constants a and b define the direction towards the light source (i.e., the sun) [HO70]. Linear FOPDE's are special cases of quasi-linear FOPDE's (in which a and b are functions of x, y and z). However, an image irradiance equation cannot be a quasi-linear FOPDE unless it is a linear FOPDE. Furthermore we can show the following lemma:

Lemma: The solutions to an image irradiance equation of the form:

$$f(ap + bq) = E(x, y) \tag{III.2.5}$$

where f is a bijection, $f^{-1}(E(x, y))$ is C^1 and where a and b are constants, can be obtained by a simple coordinate transformation from the solutions to an image irradiance equation of the form:

$$p = E(x, y). \tag{III.2.6}$$

Remark: A bijection is a one-to-one and onto function. In the above lemma, f^{-1} denotes the inverse function of f.

Proof: Since the function f is a bijection, equation (III.2.5) and the transformed equation (III.2.7):

$$ap + bq = f^{-1}(E(x, y)) \tag{III.2.7}$$

have the same solutions. We abbreviate $f^{-1}(E(x, y))$ by $\tilde{E}(x, y)$. To prove the lemma we distinguish three cases depending upon the values of a and b.

Case 1) $a \neq 0$ and $b \neq 0$

The image irradiance equation is of the form:

$$ap + bq = \tilde{E}(x, y). \tag{III.2.8}$$

Now let $z = z(x, y)$ be a solution to:

$$p = \tilde{E}(\tilde{x}, \tilde{y}) \tag{III.2.9}$$

where $\tilde{x}$ and $\tilde{y}$ are defined by:

$$\begin{aligned} \tilde{x} &= a(x+y) \\ \tilde{y} &= b(x-y). \end{aligned} \tag{III.2.10}$$

Then $\tilde{z}(\tilde{x},\tilde{y}) = z(x(\tilde{x},\tilde{y}), y(\tilde{x},\tilde{y}))$ is a solution to (III.2.8).

First we express x and y in terms of $\tilde{x}$ and $\tilde{y}$:

$$\begin{aligned} x &= \frac{\tilde{x}}{2a} + \frac{\tilde{y}}{2b} \\ y &= \frac{\tilde{x}}{2a} - \frac{\tilde{y}}{2b}. \end{aligned} \tag{III.2.11}$$

Now we determine the first order partial derivatives of $\tilde{z}$ with respect to $\tilde{x}$ and $\tilde{y}$ in terms of the first order partial derivatives of z:

$$\begin{aligned} \frac{\partial \tilde{z}}{\partial \tilde{x}} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial \tilde{x}} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial \tilde{x}} = \frac{\partial z}{\partial x}\frac{1}{2a} + \frac{\partial z}{\partial y}\frac{1}{2a} \\ \frac{\partial \tilde{z}}{\partial \tilde{y}} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial \tilde{y}} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial \tilde{y}} = \frac{\partial z}{\partial x}\frac{1}{2b} - \frac{\partial z}{\partial y}\frac{1}{2b} \end{aligned} \tag{III.2.12}$$

$$a\frac{\partial \tilde{z}}{\partial \tilde{x}} + b\frac{\partial \tilde{z}}{\partial \tilde{y}} = \frac{\partial z}{\partial x} = \tilde{E}(\tilde{x},\tilde{y}) \tag{III.2.13}$$

Thus:

$$a\frac{\partial \tilde{z}}{\partial \tilde{x}} + b\frac{\partial \tilde{z}}{\partial \tilde{y}} = \tilde{E}(\tilde{x},\tilde{y}) \tag{III.2.14}$$

which is the same equation as (III.2.8).

Case 2) $a \neq 0$ and $b = 0$

The image irradiance equation is of the form:

$$ap = \tilde{E}(x,y). \tag{III.2.15}$$

As $a \neq 0$, we can write this equation equivalently as:

$$p = \frac{\tilde{E}(x,y)}{a} \tag{III.2.16}$$

which is of the form (III.2.6).

Case 3) $a = 0$ and $b \neq 0$

The image irradiance equation is of the form:

$$bq = \tilde{E}(x, y). \tag{III.2.17}$$

As $b \neq 0$, we can write this equation equivalently as:

$$q = \frac{\tilde{E}(x, y)}{b}. \tag{III.2.18}$$

Now let $z = z(x, y)$ be a solution to:

$$p = \frac{\tilde{E}(\tilde{x}, \tilde{y})}{b} \tag{III.2.19}$$

where $\tilde{x}$ and $\tilde{y}$ are defined by:

$$\begin{aligned} \tilde{x} &= y \\ \tilde{y} &= x. \end{aligned} \tag{III.2.20}$$

Then $\tilde{z}(\tilde{x}, \tilde{y}) = z(x(\tilde{y}), y(\tilde{x}))$ is a solution to (III.2.17).

Now we show that the first order partial derivative of $\tilde{z}$ with respect to $\tilde{y}$ is the first order partial derivative of z with respect to x:

$$\frac{\partial \tilde{z}}{\partial \tilde{y}} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial \tilde{y}} = p \tag{III.2.21}$$

Thus:

$$\frac{\partial \tilde{z}}{\partial \tilde{y}} = \frac{\tilde{E}(\tilde{x}, \tilde{y})}{b} \tag{III.2.22}$$

which is the same as equation (III.2.17).

Hence it suffices to examine linear image irradiance equations of the form (III.2.6). ∎

As previously mentioned, a FOPDE has, in general, infinitely many solutions. What kinds of conditions can one impose such that the solutions to a FOPDE are restricted to a single one? A solution to the Cauchy problem which is the problem of constructing an integral surface passing through any given curve C, provides us with one answer to this question and is stated in the following theorem:

Theorem: Let $p = E(x, y)$ a linear image irradiance equation and let C be an initial, continuous and non-characteristic curve. If Δ (III.2.3) does not vanish along C, then there exists a unique, smooth integral surface through C.

Proof: The proof follows from the existence and uniqueness theorem for ordinary differential equations and can be found in [COHI62b, pp.145–147] and in section A.6. ∎

If Δ vanishes along C, then C is either a characteristic curve and the equation has infinitely many solutions, or the integral surface does not have continuous derivatives along C. In particular at the end of section A.6 the following lemma is shown:

> **Lemma**: Let $ap + bq = E(x, y)$ be a linear FOPDE. Let C be a non-characteristic initial curve for which Δ vanishes (III.2.3). Then the solutions to this FOPDE are cylindrical surfaces perpendicular to the x-y plane.

Proof: The proof can be found in section A.6. ∎

In the previous theorem C was assumed to be a continuous curve. However, it can be shown that "any singularities of the initial data propagate in the x-y plane along the projection there of the relevant characteristic curve" [GAR64, p.22]. This is not surprising since characteristic curves can be viewed as branch curves along which two integral surfaces meet.

Let $x = x(t)$ and $y = y(t)$ denote the parametric representation of a curve in the x-y plane. A point $P = (x_0, y_0)$ is called a *double point*, if several values of t correspond to P. In the case where an initial curve C or its projection onto the x-y plane has double points, the integral surface containing C has self-intersections and therefore z is not a single-valued function of x and y.

The projection of a characteristic curve onto the x-y plane is called a *base characteristic*. In the case of a linear image irradiance equation the base characteristics are unique. An integral surface of a linear continuous image irradiance equation can have a bounding contour and we can show the following lemma:

> **Lemma**: Let $ap + bq = E(x, y)$ be a continuous image irradiance equation and let an integral surface which has a bounding contour be defined by $z = z(x, y)$. Then its b-silhouette is a base characteristic.

Proof: For points on the bounding contour p and/or q are singular. Thus any initial curve which has a point in common with the bounding contour is discontinuous and the singularity propagates along a characteristic curve. ∎

III.2.2. The General Continuous Image Irradiance Equation

We discuss now general continuous image irradiance equations. A more detailed exposition of this material can be found in section A.3. In deriving the system of

characteristic equations (III.2.2) it is also assumed that an integral surface z has continuous second order partial derivatives and therefore that the equation $p_y = q_x$ holds. In [PLI54] it is shown that an integral surface can also be built from characteristic strips in the case when only its first order partial derivatives are continuous.

The Cauchy problem for a general FOPDE is formulated in the same way as for linear FOPDE's and is discussed in more detail section A.7. Let C be an initial curve specified in parametric form by $x = x(t), y = y(t)$ and $z = z(t)$. Then $p(t)$ and $q(t)$ along C can be determined by solving the two equations:

$$R(p(t), q(t)) = E(x(t), y(t)) \tag{III.2.23}$$

$$\frac{dz}{dt} = p(t)\frac{dx}{dt} + q(t)\frac{dy}{dt}. \tag{III.2.24}$$

As (III.2.23) is, in general, a nonlinear equation in p and q, several solutions may be possible for $p(t)$ and $q(t)$ along C. To "avoid inessential reference to possible multiple valuedness of solutions for p and q along C" [COHI62b, p.80] it is assumed that p and q are also known as initial data. For the following discussion we assume that p and q are specified along C. In other words the initial data is an initial strip (which we denote by C_1). Now if $\Delta \neq 0$ (III.2.3) along C then a unique integral surface may be obtained:

Theorem: Let $R(p, q) = E(x, y)$ be a continuous image irradiance equation. Then for every initial strip which is not a characteristic strip and for which $\Delta \neq 0$, there exists a unique integral surface through this strip.

Proof: The proof follows from the existence and uniqueness theorem for ordinary differential equations and can be found in [COHI62b, pp.79–82] and in section A.7. It has been shown [HA28] that if the initial data does not have continuous second order derivatives, then the solution does not have continuous first order derivatives. ∎

An integral surface of an image irradiance equation is a solution to a Cauchy problem only under the assumption that $R_p^2 + R_q^2 \neq 0$ (section A.3). Let S be an integral surface of an image irradiance equation for which the condition $R_p^2 + R_q^2 \neq 0$ holds. Then there exists an initial strip C_1 such that the solution to the characteristic equations with C_1 as initial values is the surface S. Thus the points of a reflectance map at which $R_p^2 + R_q^2 = 0$ have to be investigated separately. Such points are called *stationary points* and are defined as:

Definition: A smooth function $f(x, y)$ has a *stationary point* at (x_0, y_0) if

$$f_x(x_0, y_0) = 0 \tag{III.2.25}$$

$$f_y(x_0, y_0) = 0.$$

We discuss now the conditions under which stationary points constrain the integral surfaces of an image irradiance equation. First we must define *critical elements* of the characteristic equations of an image irradiance equation:

> **Definition**: Let $R(p,q) = E(x,y)$ be an image irradiance equation. Then a point denoted by (x_0, y_0, p_0, q_0) is a *critical element* if (x_0, y_0) is a stationary point of $E(x,y)$ and (p_0, q_0) is a stationary point of $R(p,q)$.
>
> The point (x_0, y_0, p_0, q_0) is a *critical point* of the image irradiance equation if it is a critical element and if the values (x_0, y_0, p_0, q_0) satisfy the image irradiance equation.
>
> The point (x_0, y_0, p_0, q_0) is a *singular point* if it is a a critical point for which the values (p_0, q_0) are uniquely determined by the values (x_0, y_0).
>
> A point P is an *isolated* critical element if in some neighborhood of it, it is the *only stationary point of* $E(x,y)$ and $R(p,q)$. Isolated critical (singular) points can be defined similarly.

Hence at a singular point the gradient of each integral surface can be uniquely determined from the measured brightness. For example, *the lines* $x = 0$ and $p = 0$ consist entirely of critical points of the image irradiance equation $p^2 = x^2$. Note that these points are not singular as the value for q cannot be uniquely determined at a point $(0, y)$ in the x-y plane. The point $(x, y, p, q) = (0,0,0,0)$, however, is a singular point of the image irradiance equation $p^2 + q^2 = x^2 + y^2$. Thus the tangent plane to each integral surface of this equation at the point $(0, 0, z)$ is parallel to the x-y plane. In chapter IV we will show how singular points constrain the integral surfaces of an image irradiance equation.

Discussing the Cauchy problem further, the case $\Delta = 0$ for a general FOPDE is analogous to the case $\Delta = 0$ for a linear FOPDE. *If an initial strip is a characteristic* strip, then the equation possesses infinitely many solutions. Again, the question arises as to whether it is possible to specify an initial strip C_1, (i.e., a curve C and p and q along it), such that $\Delta = 0$ and C_1 is not a characteristic strip. Such a strip can be specified but "then there exists no integral surface which contains this initial strip and has continuous derivatives up to the second order in its neighborhood" [COHI62b, p.83]. However, it might be possible to construct a surface through C_1. Then the curve C is a *singular curve* for that surface [COHI62b, p.83], i.e., along this curve the function defining the surface does not have continuous second order partial derivatives.

Thus the solutions to a continuous image irradiance equation do not necessarily have continuous partial derivatives everywhere. In particular p and q can be singular as shown in the example below. So, one integral surface of a continuous image irradiance equation can have a bounding contour although this does not imply that every integral surface of this equation has a bounding contour. For instance:

$$z(x,y) = 4x^{1/4} + f(y) \tag{III.2.26}$$

where f is any continuous function, defines an integral surface of the following image irradiance equation:

$$\frac{(p^2-1)^2}{(p^2+1)^2} = \frac{(1-x^{\frac{3}{2}})^2}{(1+x^{\frac{3}{2}})^2}. \tag{III.2.27}$$

Note that for $x = 0$, p, the first order partial derivative of z with respect to x as defined by (III.2.26), is singular. Thus this surface has a bounding contour and its image contains a b-silhouette. However $E(x,y)$ and $\frac{\partial E}{\partial x}$ are continuous along the curve $x = 0$.

On the other hand, there exist integral surfaces of (III.2.27) which do not have a bounding contour. For example, the integral surface defined by:

$$z(x,y) = \frac{4}{7}x^{\frac{7}{4}} + g(y) \tag{III.2.28}$$

where g is any continuous function is a solution to (III.2.27).

There are, nevertheless, continuous image irradiance equations for which all integral surfaces have a bounding contour as shown in the next lemma. Any such equation can always be transformed into a singular image irradiance equation such that the original and the transformed equations have the same solutions:

> **Lemma**: Let $R(p,q) = E(x,y)$ be a continuous image irradiance equation. Then all of its integral surfaces have a bounding contour, if there exist a C^1 bijection g, a C^1 function f and if the following conditions hold:
>
> - The reflectance map can be written as:
>
> $$R(p,q) = g(f(p,q)). \tag{III.2.29}$$
>
> - There exist finite values for x and y denoted by x_0 and y_0 such that:
>
> $$\lim_{\substack{x \to x_0 \\ y \to y_0}} g^{-1}(E(x,y)) = \pm\infty. \tag{III.2.30}$$

Proof: The solutions to the continuous image irradiance equation $R(p,q) = E(x,y)$ and the equation $f(p,q) = g^{-1}(E(x,y))$ are the same. Note that the points (x_0, y_0) constitute the b-silhouette. Thus by virtue of the preceding discussion of bounding contours and b-silhouettes, the lemma is self-evident. ∎

In section III.3 and chapter V we will discuss the solution to the reconstruction problem in the case where a b-silhouette can be uniquely identified in an image.

The last case of interest here is where an initial curve C degenerates to a point P (see also section A.7 and note that the following discussion is relevant only for FOPDE's

which are not quasi-linear). Then "all characteristic curves through a fixed point P of x, y, z-space form an integral surface" [COHI62b, p.83]. Such a surface S has a conical singularity at P and the Monge cone is the tangent cone to S at P. A surface constructed in such a manner is called the *integral conoid* of the partial differential equation. For a given image irradiance equation and a given point P there exists exactly one integral conoid.

In summary, the method of characteristic curves can be used to determine the solutions to a FOPDE. An initial curve which lies on the integral surface must, in general, be specified in order to restrict the solutions to a continuous image irradiance equation to a single one. The critical points of an image irradiance equation have to be analyzed separately.

III.2.3. Edges and Vertices

In this section we give examples of how to formulate the Cauchy problem such that its solution is an integral surface containing an edge. As previously stated, if an initial curve C is discontinuous then the surface normals to the integral surface which contains C are discontinuous. We give a very simple example illustrating this. Two planes intersect along a straight line as depicted in figure 2. The image irradiance equation is assumed to be linear and the same for both planes:

$$p + q = 1. \tag{III.2.31}$$

We pose the following questions: Do the planes have to be oriented in a particular way to assure that the previous image irradiance equation holds? Are there any other integral surfaces which contain the edge constituted by the intersection of two such planes? The answers to these questions depend upon additional information about the scene which is available. Referring to figure 2, if we specify the curve K in three space, it can be used as initial data to uniquely reconstruct the two planes from the image irradiance equation. As the image irradiance equation is linear, the curve T has to be a characteristic. We observed earlier (section III.2.1) that characteristic curves can be viewed as branch curves at which two different integral surfaces meet. Through every curve which is not a characteristic there exists exactly one smooth surface. Thus if an integral surface of (III.2.31) contains an invisible edge, its projection onto the x-y plane is a line whose slope is 1. (The characteristic curves of a linear image irradiance equation whose reflectance map is of the form $R(p, q) = ap + bq$ where a and b are nonzero constants, are straight lines whose slope is proportional to $\frac{b}{a}$.) Recall that the linear reflectance map describes, for instance, the reflectivity properties of the maria of the moon and that the constants a and b specify then the direction towards the sun (section III.2.1). We conclude therefore that any ridges on the moon which are not

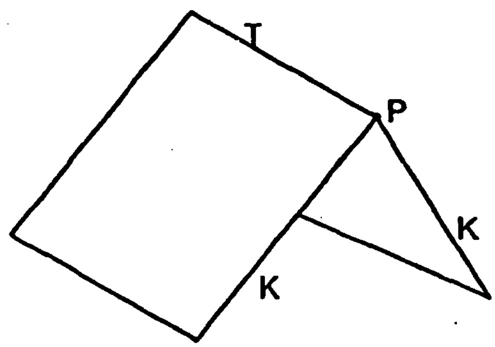


Figure 2. Two intersecting planes

visible are in the direction of the sun.

On the other hand, if only the edge T is specified, the two surfaces (which may not be planes) cannot be reconstructed in a unique way from a linear image irradiance equation. This follows from the fact that T has to be a characteristic curve. For example, the surfaces defined by the next two equations are solutions to (III.2.31):

$$z(x,y) = x + (y-x)^2 \tag{III.2.32}$$

$$z(x,y) = x + \sin(y-x). \tag{III.2.33}$$

At their intersection these two surfaces form an edge which is defined by:

$$x = y \qquad\qquad z = x. \tag{III.2.34}$$

Furthermore the two planes (which are also solutions to (III.2.31)) defined by:

$$x - 2y + z = 0 \tag{III.2.35}$$

$$2x - 3y + z = 0 \tag{III.2.36}$$

give rise to the same edge.

If an image irradiance equation is nonlinear, however, there is only a small number of surfaces which can give rise to a specific edge. In particular an edge cannot be a characteristic curve, as two integral surfaces which meet along a characteristic curve also have the same tangent plane there. So let T be an edge given in parametric form. Using equations (III.2.23) and (III.2.24) we can determine the possible values for p and q along T. Since the image irradiance equation is assumed to be nonlinear, several solutions for p and q are possible, so many different integral surfaces can be constructed

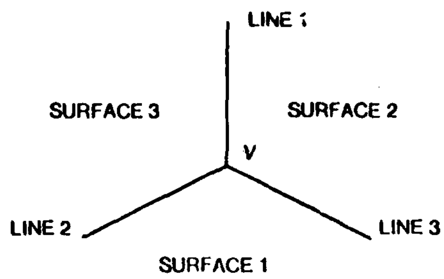


Figure 3. Vertex

each of which has the edge embedded in it. Any two of these surfaces will intersect at an edge. Hence in the case of a nonlinear image irradiance equation, only a small number of different surfaces can give rise to a particular edge.

Let us now specify a vertex V in space from which three straight lines emanate (referred to as a *corner*) as shown in figure 3. The equations of the three lines are given in three space. Assume, as well, that the following image irradiance equation holds in all three regions:

$$p^2 + q^2 = 1. \tag{III.2.37}$$

We ask how many integral surfaces exist which have the same shading and contain the vertex and the three lines. (Note that these lines are not visible.) Clearly, all such surfaces will have a *singularity* at the vertex, i.e., the function $z = z(x, y)$ defining a surface is not differentiable at V. A priori there are two ways of interpreting the lines leading from the vertex:

1) as curves lying on a smooth surface or

2) as edges.

We will discuss both interpretations.

Assuming case 1, we are faced with the problem of constructing a smooth (except at V) surface S which has a conical singularity at V and has the three lines embedded in it. Thus the vertex is a *degenerate* initial curve (see sections III.2.2 and A.7) and S is the integral conoid. If such a surface S exists, the image irradiance equation cannot be linear, the generators of the integral conoid must be straight lines and the vertex

cannot be a singular point of the image irradiance equation. In section III.2.2 it was shown that there cannot be another surface besides the integral conoid which has a conical singularity at V as the integral conoid is uniquely determined by the image irradiance equation and this vertex. So if the integral conoid with its vertex at V can be constructed such that the three lines lie on it, the surface S is unique. In the case where V is a singular point or the generators of the integral conoid are not straight lines, no surface S can be found and the three lines necessarily specify three edges.

We show now that to be able to construct the surface S as described in the preceding paragraph, the three lines necessarily are characteristic curves of the image irradiance equation. So suppose that the integral conoid can be constructed at V such that it contains the three lines. Then the integral conoid and the Monge cone are identical, which follows from the fact that the Monge cone is the tangent cone to S at V. Thus if the integral conoid has straight lines, they must coincide with the generators of the Monge cone. The lines leading from the vertex are therefore characteristic curves of the image irradiance equation. In other words, if the lines are characteristic curves and if V is not a singular point, then an integral conoid which contains the corner exists.

In our particular example the integral conoid of (III.2.37) is a right circular cone whose generators are inclined to the x-y plane at 45 degrees. (A cone is called a *right circular cone* if the angle between its axis and a plane containing the circular cross section is 90 degrees.) Thus if the three lines are characteristic curves, the right circular cone is the desired surface S.

The other interpretation of the corner is that the three lines originating at the vertex specify three edges. Again only in the case where the image irradiance equation is nonlinear can we possibly find three surfaces which form the three edges since, if the image irradiance equation is linear, different integral surfaces intersect only along characteristic curves. Yet the base characteristics are all parallel for a linear equation, hence the projection of the three lines onto the x-y plane cannot be base characteristics and so the lines cannot be characteristics.

We want then to find three surfaces which intersect along the three lines using these lines as initial data. For a particular corner, three such surfaces do not necessarily exist as can be easily seen using equation (III.2.37). The integral surfaces of (III.2.37) that are constructed using a straight line (which is not a characteristic) as initial data, are planes inclined at 45 degrees to the x-y plane. Although through three lines which constitute a corner we can always find three planes, they are not necessarily inclined at 45 degrees to the x-y plane and are therefore not necessarily integral surfaces of (III.2.37). However, if three integral surfaces do exist, they are unique. From the two lines which are embedded in every surface, p and q can be uniquely determined. If these values for p and q as functions of x and y satisfy the image irradiance equation then each of the surfaces can be uniquely determined.

Summarizing the observations made above: if three lines which form a corner are characteristic curves of an image irradiance equation, then the surface containing the corner is the integral conoid of the equation. Otherwise such a corner can be used to

reconstruct the three surfaces which form it.

III.2.4. Gaussian Image

In the previous sections we reviewed the well known results concerning continuous FOPDE's and showed that in general, a continuous FOPDE has infinitely many solutions. Only after imposing some additional constraints (in particular by specifying an initial strip which is not a characteristic strip) can the solutions be restricted to a single one.

It is clear, then, that a continuous image irradiance equation does not contain sufficient information to uniquely reconstruct the shape of a surface from its image. What *partial* knowledge about a surface does an image irradiance equation provide? To understand this problem better, certain concepts taken from differential geometry are now introduced as they provide us with a formalism to discuss the *shape* of a surface. Some technical prerequisites are first reviewed briefly. A more detailed exposition can be found in any standard book on differential geometry, e.g., [CAR76]. For the rest of this section we assume that a function $z = z(x, y)$ which defines a surface is C^2.

The *Gaussian sphere* is a sphere of unit radius. The surface normal at every point on the Gaussian sphere can have two possible orientations, referred to as the positive (or outward) and negative (or inward) directions. We adopt the convention that a surface normal on the Gaussian sphere points outward. The *Gaussian mapping* maps a point P on a surface into a corresponding point P^G on the Gaussian sphere such that P and P^G have the same surface normal. This mapping is well defined since no two points on the Gaussian sphere have the same normal. The *Gaussian image* (sometimes referred to as the spherical image) of a connected patch of a smooth surface maps into a connected region on the Gaussian sphere. The area of the Gaussian image of a surface is called its *integral curvature.*

Points on a surface are classified according to the behavior of the tangent planes with respect to the surface:

> **Definition**: A point P on a surface is *elliptic* if the tangent plane at P does not intersect the surface at any other point. A point P is *hyperbolic* if the tangent plane at P intersects the surface in a curve which has two branches intersecting at P. A point P is *parabolic* if the tangent plane at P intersects the surface along a single curve (which can degenerate to a point).

A surface which consists only of elliptic points is *locally convex*. We will say a surface is *hyperbolic* if it consists only of hyperbolic points. The points which separate regions where a surface consists of elliptic and hyperbolic points, respectively, are parabolic points.

The grouping of points on a surface into elliptic, hyperbolic and parabolic points

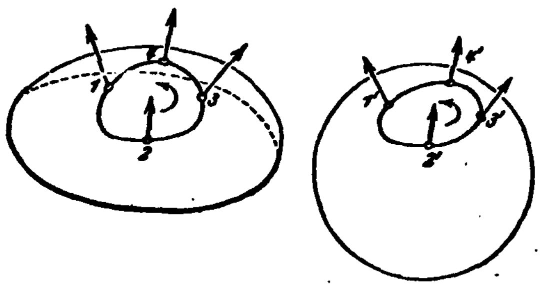


Figure 4. Gaussian image of a neighborhood of an elliptic point

can easily be expressed in terms of *Gaussian curvature*. Let F be a simply connected and bounded surface patch which *is enclosed by the closed curve* k and let G denote the Gaussian image of F, which is enclosed by the closed curve k^G. "We divide the area G enclosed by k^G on the sphere by the area F enclosed by k on the surface and then shrink the curve k down to a point P on the surface. Then F and G approach zero and their quotient approaches a definite limit K:

$$\lim_{F \to 0} \frac{G}{F} = K. \tag{III.2.38}$$

The number K defined in this way is called the Gaussian curvature" [HICV52, pp.193-194]. A point P on a surface is elliptic if the Gaussian curvature is positive there, hyperbolic if the Gaussian curvature is negative, and parabolic if the Gaussian curvature is zero. If $z = z(x, y)$ specifies a surface and is C^2, then the Gaussian curvature at a point (x, y, z) is defined by:

$$K(x, y) = \frac{z_{xx} z_{yy} - z_{xy}^2}{(1 + z_x^2 + z_y^2)^2}. \tag{III.2.39}$$

The common notions of a surface being *convex* or *concave* do not refer to the type (as defined above) of points a surface consists of. They merely distinguish the two sides of a locally convex surface (or correspond to the two possible directions of a normal vector) with respect to a viewer.

The mapping between a surface and the Gaussian sphere is one-to-one if the surface is either locally convex or hyperbolic. "If we move around an elliptic point along a small closed curve that lies on the surface, its spherical image - assuming that the surface has no double points - will also be a closed curve without double points, and *this curve is traversed in the same sense* as the original curve (figure 4). A small curve without double points about a hyperbolic point is also mapped into a curve without

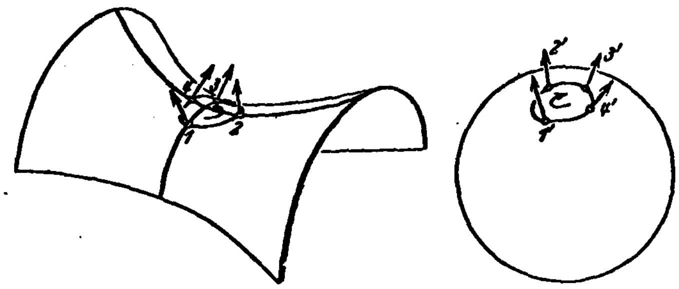


Figure 5. Gaussian image of a neighborhood of a hyperbolic point

double points, but in this case the sense is reversed (figure 5)" [HICV52, pp.195-196]. The spherical image of a surface which consists of elliptic, hyperbolic and parabolic points consists of several *sheets*, i.e., several points on a surface get mapped into the same point on the Gaussian sphere.

We will also need the definitions of a *closed* and a *compact* surface:

Definition [CAR76, p.112]: Let A be a subset of $\Re^3$. We say that $p \in \Re^3$ is a limit point of A if every neighborhood of p in $\Re^3$ contains a point of A distinct from p. A is said to be *closed* if it contains all of its limit points. A is *bounded* if it is contained in some ball of $\Re^3$. If A is closed and bounded it is called a *compact* set.

For example, the surface of a sphere is compact, whereas a paraboloid of revolution defined by $z(x,y) = x^2 + y^2$ is a closed but not compact surface.

The notion *similar* captures formally what we mean by saying that two surfaces have the same shape:

Definition: Two surfaces in $\Re^3$ are *similar* if they can be mapped into each other by a composition of translations, rotations, reflections and dilations.

III.2.5. Gradient Space

In this section we will briefly discuss gradient space as popularized by Mackworth [MAC73] and Horn [HO77] and which is now a standard tool in vision research. Our objective is to show how the concepts from differential geometry help us to understand the variety of integral surfaces of an image irradiance equation. We will also investigate how different constraints may restrict the number of possible solutions to a given image irradiance equation.

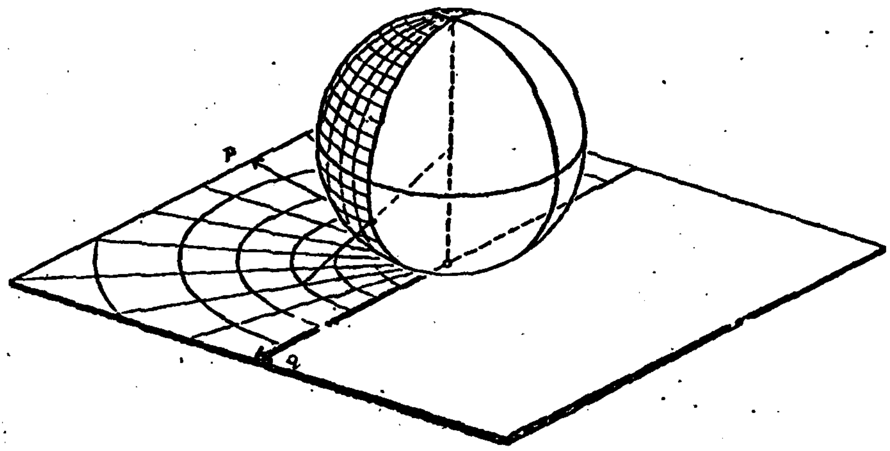


Figure 6. Map: reflectance map → Gaussian sphere

Let

$$R(p,q) = E(x,y) \tag{III.2.40}$$

be an image irradiance equation. For simplicity of exposition we assume that $E(x,y)$ is defined at every point in the x-y plane. Then we can write the previous equation as a system of two equations:

$$R(p,q) = c \tag{III.2.41}$$

$$E(x,y) = c \tag{III.2.42}$$

where c is a constant. In the p-q plane (also called *gradient space*), the graph of $R(p,q) = c$, for all possible values of c, is called the *reflectance map*. A reflectance map can be mapped onto the Gaussian sphere by placing the south pole of the Gaussian sphere onto the origin of the p-q plane by means of a simple projection from the center of the sphere, as illustrated in figure 6. Note that the mapping between gradient space and the Gaussian sphere is conformal. The graph of $E(x,y) = c$, for all possible values of c, can be drawn in the x-y plane, referred to as the image plane. For example consider the following eikonal equation:

$$p^2 + q^2 = x^2 + y^2. \tag{III.2.43}$$

This equation can be rewritten as:

$$p^2 + q^2 = c \tag{III.2.44}$$

$$x^2 + y^2 = c.$$

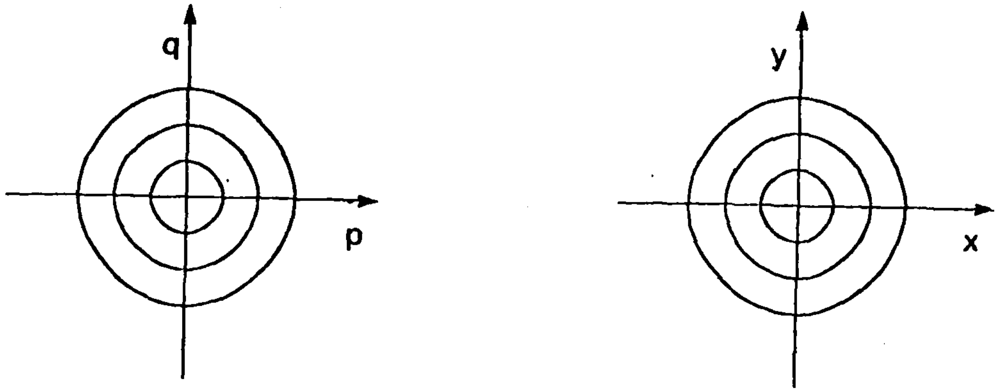


Figure 7. Graphs for the equation $p^2 + q^2 = x^2 + y^2$

The graphs for these equations are depicted in figure 7.

Let $z = z(x, y)$ define an integral surface of an image irradiance equation. Then the gradient at every point of z is the vector $(p, q, -1)$ where $p = \frac{\partial z}{\partial x}$ and $q = \frac{\partial z}{\partial y}$. As discussed in chapter II, we assume that each point $P = (x_0, y_0, z(x_0, y_0))$ on the surface is mapped via orthographic projection into the point (x_0, y_0) in the image plane. We also define a mapping from surface orientation to gradient space: The gradient $(p_0, q_0, -1)$ at P is mapped into the point (p_0, q_0) in gradient space.

An image irradiance equation gives us some information about the correspondence between points in the image plane and points in gradient space. In particular, if (x_0, y_0) lies on the curve $E(x, y) = c_0$ for some constant c_0, then (p_0, q_0) lies on the curve $R(p, q) = c_0$. Note that determining an integral surface (up to a constant factor) of an image irradiance equation is equivalent to specifying the correspondence between points in the image plane and points in gradient space.

We want to show now that any two integral surfaces of an image irradiance equation need not be similar. For two such surfaces to be similar, it is necessary that they consist of the same type (i.e., elliptic, hyperbolic or parabolic) of points. An image irradiance equation, however, does not restrict the type of points on the integral surfaces. So by knowing one integral surface of an image irradiance equation, other such surfaces cannot, in general, be obtained through similarity transformations. To prove these assertions we examine some integral surfaces of (III.2.43). In particular the surface defined by the following equation consists entirely of elliptic points:

$$z(x, y) = \frac{x^2 + y^2}{2} \qquad \text{(III.2.45)}$$

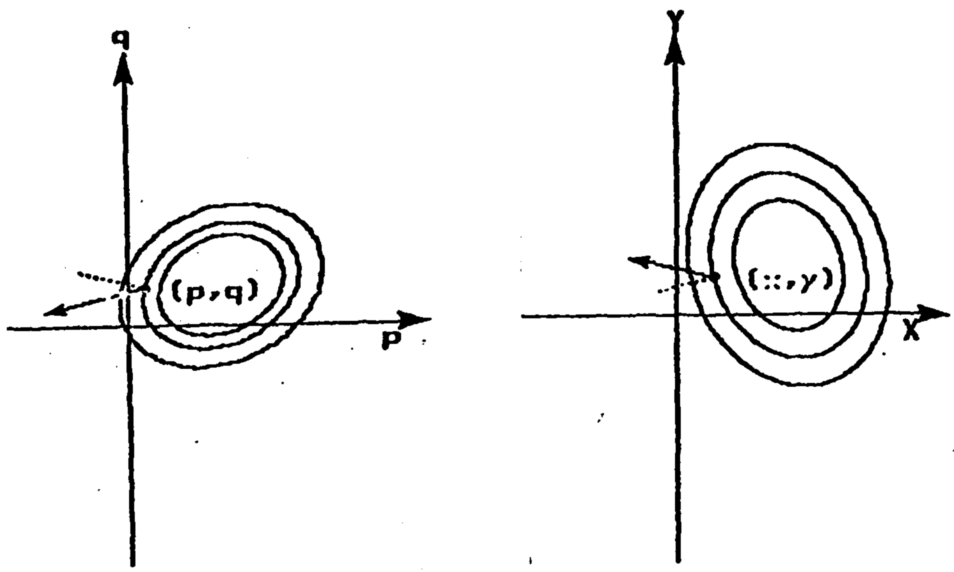


Figure 8. Graph for characteristic curves

whereas the surface defined by:

$$z = xy \tag{III.2.46}$$

is hyperbolic.

As already noted several times, the basic tool for solving a FOPDE is the method of characteristic curves. Using gradient space, the characteristic curves of an image irradiance equation and the necessity of specifying an initial curve to restrict the possible solutions to a single one can be visualized easily [HO75]. In the following we use only four of the five characteristic equations (III.2.2) of an image irradiance equation $R(p, q) = E(x, y)$:

$$\frac{dx}{ds} = R_p(p, q) \tag{III.2.47}$$

$$\frac{dy}{ds} = R_q(p, q) \tag{III.2.48}$$

$$\frac{dp}{ds} = E_x(x, y) \tag{III.2.49}$$

$$\frac{dq}{ds} = E_y(x, y). \tag{III.2.50}$$

In the image plane, dx and dy can be viewed as the two components of a vector and E_x and E_y as the direction of a normal vector to the curve $E(x, y) = c$ (where c is a constant). In the same way dp and dq can be interpreted in gradient space as the components of a vector and R_p and R_q as the direction of a normal vector to the iso-brightness curves $R(p, q) = c$. From equations (III.2.47) and (III.2.48) one can deduce that the two vectors (dx, dy) and (R_p, R_q) are parallel. Similarly, equations (III.2.49) and (III.2.50) indicate that the two vectors (dp, dq) and (E_x, E_y) are parallel.

Now let the point (x_0, y_0) in the image plane correspond to the point (p_0, q_0) in gradient space (see also figure 8 which is taken from [WOOD78, p.190]). Then a step in the image plane from the point (x_0, y_0) in the direction of a characteristic curve, corresponds to a step in gradient space from (p_0, q_0) in the direction (E_x, E_y). Conversely, a step from the point (p_0, q_0) in the direction (dp, dq) corresponds to a movement in the image plane starting at (x_0, y_0), in the direction (R_p, R_q). Specifying an initial curve now gives a correspondence between a set of points, denoted by X, in the x-y plane and a set of points, denoted by P, in gradient space. If this initial curve is not a characteristic, the characteristic curves can be expanded from each point in X which just corresponds to expanding curves in gradient space from every point in P. In this manner a point (p, q) is assigned to each point in the image plane. Once p and q are known at every point on the surface, the function $z = z(x, y)$ defining the surface can be found by integration (assuming that p and q satisfy the equation $p_y = q_x$). Thus to determine a unique (up to translation in the z-direction) solution to an image irradiance equation, enough information has to be known so that the characteristic curves can be expanded simultaneously in the image plane and in gradient space.

III.3. The Singular Image Irradiance Equation

In section III.1 we classified image irradiance equations into two categories. In the previous sections, we discussed equations which fall into the first category, i.e., the case where $E(x, y)$ is C^1. This section deals with the second class, i.e., the case where $E(x, y)$ is ***singular***.

Recall (chapter I) that a function $E(x, y)$ is singular at a point (x_0, y_0) if:

$$\lim_{\substack{x \to x_0 \\ y \to y_0}} E(x, y) = \pm\infty. \tag{III.3.1}$$

Lemma: Let $R(p, q) = E(x, y)$ be a singular image irradiance equation. Then the b-silhouette consists of those points (x_0, y_0) in the x-y plane for which $E(x, y)$ is singular.

Proof: Each singular image irradiance equation defines a b-silhouette. The lemma follows. ∎

Let us recall here briefly the geometry of the image forming system as discussed in chapter II. It is assumed that the viewing direction coincides with the z-axis. Furthermore, we approximate the imaging situation by orthographic projection hence the lines connecting the viewer and points on the surface are parallel to each other and perpendicular to the x-y plane. For simplicity we will restrict our attention to images where a single b-silhouette occurs. Images with multiple b-silhouettes can be treated in a similar fashion. We will call a b-silhouette *nondegenerate* if it is a connected and smooth curve in the x-y plane.

Let $R(p,q) = E(x,y)$ be a singular image irradiance equation, let the b-silhouette be defined by $w(x,y) = 0$ and let $z = z(x,y)$ define an integral surface. Then its bounding contour consists of the points $(x_0, y_0, z(x_0, y_0))$ on the integral surface such that (x_0, y_0) belongs to the b-silhouette. We assume that integral surfaces are (piecewise) smooth (section III.1). So, no parts of an integral surface obscure each other and therefore the bounding contour is a set of (piecewise) continuous curves. As previously stated, the lines connecting the viewer and points on the bounding contour graze the surface. In other words, at every point on the bounding contour the tangent plane is perpendicular to the x-y plane. In terms of p and q this means that p and/or q assume infinite value at points on the bounding contour. The spherical image of points on the bounding contour (section III.2.4) corresponds to points on the equator of the Gaussian sphere.

If the equation of the b-silhouette is known, the surface normal to points on the bounding contour can be determined. (This follows directly from the definition of bounding contour. The surface normal at a point $(x_0, y_0, z(x_0, y_0))$ on the bounding contour is $(\pm w_x(x_0, y_0), \pm w_y(x_0, y_0), 0)$ where the $+$ and $-$ sign distinguish the two sides of a surface with respect to the viewer.) This surface is tangent to a cylinder whose intersection with the x-y plane is the b-silhouette and whose generators are perpendicular to the x-y plane.

So again we pose the question: What are the constraints necessary to obtain a unique solution to a singular image irradiance equation? We shall proceed in two ways. In section III.3.2 we show that the method of characteristic curves can be used to find the integral surfaces of a singular image irradiance equation. Unfortunately, specifying a bounding contour does not restrict the possible integral surfaces to a single one, as shown by means of an example in section III.3.2. In chapter V we investigate constraints which enable us to solve the reconstruction problem uniquely.

The existence of a smooth integral surface is not guaranteed for every singular image irradiance equation as we now demonstrate. It is important to notice that we are looking for a global solution to an image irradiance equation, i.e., an integral surface which is defined at every point (x,y) for which the equation is defined. In particular, an integral surface should be bounded for points (x,y) which lie on the b-silhouette. An example of an image irradiance equation for which no global bounded solution exists is given by:

$$p = \frac{1}{x}. \tag{III.3.2}$$

The general solution to this equation is:

$$z(x, y) = \ln x + f(y) + c \tag{III.3.3}$$

where f is any continuous function and c is a constant. For points on the y-axis (i.e., $x = 0$) $z(x, y)$ as defined by the previous equation is not bounded.

For any given singular image irradiance equation, the points on the b-silhouette can be found by inspection of $E(x, y)$. The following example should clarify this:

$$p^2 + q^2 = \frac{x^2 + y^2}{1 - (x^2 + y^2)}. \tag{III.3.4}$$

Along the circle $x^2 + y^2 = 1$, $E(x, y)$ assumes infinite value. An integral surface of the previous equation is a sphere defined by:

$$z(x, y) = \sqrt{1 - (x^2 + y^2)} \tag{III.3.5}$$

with p and q given by:

$$\begin{aligned} p &= \frac{-x}{\sqrt{1 - (x^2 + y^2)}} \\ q &= \frac{-y}{\sqrt{1 - (x^2 + y^2)}}. \end{aligned} \tag{III.3.6}$$

Along the circle $x^2 + y^2 = 1$, p and q are infinite but the surface normal there is well defined as $(x, y, 0)$. Note also that if $z = z(x, y)$ is a solution to an image irradiance equation, then so is $\tilde{z} = z(x, y) + c$ for any constant c. As we are solely interested in determining the shape of surfaces, we will assume in the following chapters that c is given. (In other words, all integral surfaces are determined only up to translation in the z-direction. Thus when we say, for instance, that there is a *unique* integral surface of an image irradiance equation we mean that it is unique up to translation in the z-direction.) In section V.2 we show that the convex and the concave hemispheres are the only two integral surfaces of (III.3.4) which have continuous second order partial derivatives.

III.3.1. Marr's Occluding Contours

One of the first to investigate b-silhouettes was David Marr [MA77]. His nomenclature differs from ours and to avoid confusion we compare the two terminologies. In his

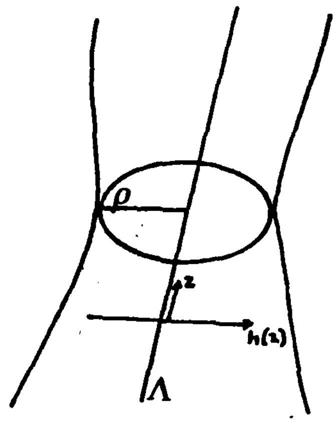


Figure 9. Generalized cone

paper bounding contours are referred to as contour generators, and b-silhouettes, as contours.

Marr poses the question of what can be inferred about the shape of a surface from its b-silhouette alone. In order to more easily approach this problem, a priori assumptions about the surface must be made. The restrictions imposed are:

R1.) The surface is defined by a C^2 function.

R2.) Each point on the bounding contour projects to a different point on the b-silhouette.

R3.) Nearby points on the b-silhouette arise from nearby points on the bounding contour.

R4.) The bounding contour is planar.

As discussed in the paper, the first three restrictions are not severe. Marr points out that for bounded surfaces, R3 follows from R2 but he nevertheless chooses to state the two restrictions separately as "they have sufficiently different meaning" [MA77]. The third restriction states that there are no gaps in the viewing direction. Implied by this is that a bounding contour cannot be created by two surfaces, one of which partially occludes the other, as this would violate R2. A consequence of the first three restrictions is that a bounding contour is a continuous curve.

The main technique used in [MA77] to interpret b-silhouettes is to examine their inflexion points, i.e., those points which separate convex regions on a b-silhouette from

concave ones. If a bounding contour is assumed to be planar, inflexion points on the b-silhouette imply the existence of inflexion points on the bounding contour. This observation leads to restriction R4 which as Marr observes, is a very strong one. He is therefore only able to analyze a small number of scenes.

The main theorem of [MA77] (theorem 1) states that generalized cones are the only type of surfaces which satisfy the four restrictions above plus an additional assumption about the viewing direction. "A generalized cone (as shown in figure 9) is the surface generated by moving a smooth cross section ρ along a straight axis Λ. The cross section may vary smoothly in size (as prescribed by the axial scaling function $h(z)$), but its shape remains constant" [MA77]. Not only is R4 a very strong assumption, but for theorem 1 to hold, the viewing direction is confined to a plane which lies parallel to a cross section of the cone whose shape we wish to determine. However, a priori, there is no way to determine if the viewing direction satisfies the constraint of the theorem!

The proof of theorem 2 as stated in [MA77] is wrong. This theorem claims that a surface obeys the four previously stated restrictions for *all* viewing directions if and only if it is a quadratic surface. In the proof given it is assumed that the surface can be defined by a polynomial; the theorem is therefore only shown to hold for this special case.

III.3.2. The Method of Characteristic Curves

Let us recall here that the method of characteristic curves can be used to solve a continuous image irradiance equation. Is this method still valid in the case where an equation is singular? In fact, the answer is positive. The method of characteristic curves is based on a local theory as stated in [COHI62b, p.62]: "It should be emphasized again that all statements and derivations are *in the small*, i.e., they concern merely neighborhoods of points, etc., without necessarily specifying the extension of these neighborhoods." It is precisely the local nature of this theory which allows us to apply it to singular image irradiance equations.

Let $R(p,q) = E(x,y)$ be a fixed, singular image irradiance equation. The characteristic curves can be constructed in a fashion analogous to the continuous case. A difficulty arises only when the Cauchy problem is stated for this equation. In the continuous case, for any strip denoted by C_1 (as defined in section III.2.2) which is not a characteristic strip and for which $\Delta \neq 0$ (III.3.3), an integral surface can be found which has C_1 embedded in it. Furthermore this integral surface is uniquely specified by the FOPDE and the strip C_1. The theorem used to obtain this uniqueness result is the existence and uniqueness theorem for ordinary differential equations. Unfortunately, in some neighborhood of a b-silhouette this theorem does not hold anymore, so there is no guarantee that the equation will have a solution for *any* initial conditions.

To understand why the existence and uniqueness theorem for ordinary differential

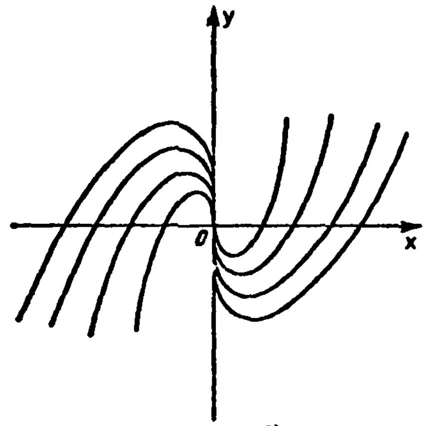


Figure 10. Solutions to III.3.8

equations does not hold at a singularity, we examine the equation:

$$\frac{dy}{dx} = \frac{x+y}{x}. \tag{III.3.7}$$

which is singular at $(x, y) = (0, 0)$. The set of solutions are:

$$y = x \ln(cx) \tag{III.3.8}$$

where c denotes any constant. The graphs of some of the curves defined by the previous equation are depicted in figure 10 [SMI68a, p.148].

The initial condition that the solution must pass through the point P, which has the coordinates $(0, 0)$, does not suffice to pin down the solution to (III.3.7) uniquely. Furthermore there is no solution which passes through the point P^0 with the coordinates $(0, 3)$. This implies that for the initial condition P^0 no solution exists. In summary, one cannot claim that for *any* initial condition, (III.3.7) can be solved in a unique way. Moreover, the type of singularity in an ordinary differential equation constrains what initial data is consistent.

If an image irradiance equation is singular, i.e., if $E(x, y)$ is singular, then some or all of its characteristic equations are singular. Thus when specifying a bounding contour, there is no guarantee that a solution exists which has the given bounding contour embedded in it. Furthermore we can show the following lemma:

Lemma: A singular image irradiance equation can have several integral surfaces each of which has the same bounding contour.

Proof: We will prove the lemma by showing it for a particular image irradiance equation. The two surfaces defined by (III.3.10) are solutions to (III.3.9) and each has the same bounding contour:

$$p^2 + q^2 = \frac{1}{x^{\frac{3}{2}}} + 1 \tag{III.3.9}$$

$$\begin{aligned} z(x,y) &= -4x^{\frac{1}{4}} + y \\ z(x,y) &= 4x^{\frac{1}{4}} + y. \end{aligned} \tag{III.3.10}$$

∎

In chapter V we shall prove that the only integral surfaces of (III.3.4) are the convex and concave hemispheres. So for any bounding contour C other than the one specified by $x^2 + y^2 = 1, z = 0$, no integral surface exists which has C embedded in it. Note in addition that the concave and the convex hemispheres have the same bounding contour. By assuming that the surface is convex in the direction of the viewer, the solutions to (III.3.4) are restricted to a single one although this is not true if we forego this assumption, as shown in the proof of the previous lemma.

Chapter IV

Singular Points

IV.1. Basic Concepts

The question which we analyze in this chapter is: How much information about the shape of a surface can one obtain from a singular point of an eikonal equation? At a singular point of an image irradiance equation the gradient of all its integral surfaces is defined by the image intensities there (section III.2.2). As discussed in chapter I, equations of the form:

$$p^2 + q^2 = E(x, y) \tag{IV.1.1}$$

are called *eikonal* equations and describe for instance the flux of the secondary electrons in a scanning electron microscope since this varies approximately as $f(p^2 + q^2)$ where f is a continuous function [LAWH60].

To obtain our results we will impose some technical conditions upon $E(x, y)$ (which we will discuss later in this section) and shall refer to eikonal equations which satisfy these conditions as *constrained* eikonal equations. The two results which we prove are:

- There exist exactly two locally convex integral surfaces of a constrained eikonal equation in some neighborhood of a singular point.

- At a singular point, the Gaussian curvature of each integral surface of a constrained eikonal equation has the same absolute value.

The first statement can be expressed in other words as: if $z = z(x, y)$ defines one locally convex solution, then $\tilde{z} = -z(x, y)$ defines the other. Hence this result can be

viewed as a uniqueness result modulo the concave/convex ambiguity. The second result exhibits a situation where a "property" of surfaces can be determined from their given (common) image. Our theorems apply also to more general equations: in appendix II we exhibit a class of image irradiance equations whose solutions can be obtained by solving an appropriate eikonal equation.

To characterize a constrained eikonal equation we need the following definition:

Definition: A function $g(x,y)$ *vanishes precisely to second order* at the point $(0,0)$ if its (limited) Taylor series expansion at $(0,0)$ is of the form:

$$g(x,y) = \alpha x^2 + \beta xy + \gamma y^2 + o((|x| + |y|)^2) \tag{IV.1.2}$$

where α, β and γ are constants, at least one of which is nonzero.

Then a constrained eikonal equation is defined as:

Definition: An eikonal equation $p^2 + q^2 = E(x,y)$ is *constrained* if $E(x,y)$ is a C^3 function satisfying the following conditions in some neighborhood of the point (x_0, y_0):

$$\begin{aligned} &1) \quad (x_0, y_0) \quad \text{is a stationary point of} \quad E(x,y) \\ &2) \quad E(x_0, y_0) = 0 \\ &3) \quad E(x,y) > 0 \quad \text{for} \quad (x,y) \neq (x_0, y_0) \\ &4) \quad E(x,y) \quad \text{vanishes precisely to second order at} \quad (x_0, y_0). \end{aligned} \tag{IV.1.3}$$

Let us discuss these conditions a bit further. The reflectance map of an eikonal equation is $R(p,q) = p^2 + q^2$ and therefore its stationary point is given by $p = 0$ and $q = 0$. We have imposed the condition that (x_0, y_0) be a stationary point of $E(x,y)$. Thus the point $P = (x, y, p, q) = (x_0, y_0, 0, 0)$ is a critical point of a constrained eikonal equation. Whence it follows from condition 2 that P is a singular point of such an equation. From the third condition we can deduce now that P is an *isolated* singular point (section III.2.2). By using a suitable linear transformation we may assume, without loss of generality, that the point $(0,0)$ is the stationary point of $E(x,y)$. Since $E(x,y)$ is assumed to be positive near the origin:

$$\alpha x^2 + \beta xy + \gamma y^2 > 0 \quad \text{for} \quad (x,y) \neq (0,0) \tag{IV.1.4}$$

defines a positive bilinear form. Thus the subsequent inequality [BRSE75, p.182] holds:

$$\alpha\gamma - \frac{\beta^2}{4} > 0. \tag{IV.1.5}$$

Moreover, the constants α and γ in the (limited) Taylor series expansion of $E(x,y)$ must be positive.

As mentioned earlier, if $z = z(x,y)$ defines a locally convex solution, then so does $\tilde{z} = -z(x,y)$. Hence, to simplify subsequent discussions, we will say that $z(x,y)$ is a locally convex *solution* to a constrained eikonal equation if it satisfies the following *positivity* conditions in some neighborhood of the origin:

$$\begin{aligned} &1) \quad z(0,0) = 0 \\ &2) \quad z(x,y) \in C^2 \\ &3) \quad z(x,y) \geq 0. \end{aligned} \tag{IV.1.6}$$

Throughout this chapter it is assumed that a solution $z = z(x,y)$ satisfies the positivity conditions and that the origin is the isolated singular point of a constrained eikonal equation.

Horn [HO75] used singular points to compute initial conditions sufficient to solve an image irradiance equation. In particular he assumed that an integral surface is convex at a singular point. In general, however, such a surface does not exist. Our results show when Horn's method can be used and we prove that in those cases, no initial conditions are needed to compute the convex surface.

IV.2. Preliminaries

The results we prove in this chapter require a number of technical prerequisites which are introduced in this section. One of the key concepts is that of a *Taylor series*, which is discussed in order to deal with the problem of approximating a function by polynomials. To be able to write the Taylor series expansion of a C^k function $z(x,y)$ in a concise form we introduce the.notion of a *homogeneous polynomial*, which is a sum of terms of the same degree:

Definition: A polynomial $P(x,y)$ is a *homogeneous polynomial of degree k* if it is of the form:

$$P(x,y) = \sum_{j=0}^{k} h_j x^j y^{k-j} \tag{IV.2.1}$$

where for each j, h_j denotes a constant.

Definition: Let $z(x,y)$ be a C^k function. Then its *(limited) Taylor series expansion* at the origin is:

$$z(x,y) = z_0 + \sum_{j=1}^{k} z_j + o((|x| + |y|)^k) \tag{IV.2.2}$$

where z_0 denotes a constant and for each j, z_j is a homogeneous polynomial of degree j. For any j such that $0 < j \leq k$, there exists exactly one polynomial z_j which matches z at the origin up to the j-th derivative.

Let us denote by f the first $k+1$ terms in the Taylor series expansion of $z(x,y)$ at the origin. Then f *approximates* z up to k-th order, i.e., the error between f and z vanishes faster than any polynomial of degree k.

Now we investigate the characteristic equations of a constrained eikonal equation and their solutions in some neighborhood of an isolated singular point. The four relevant characteristic equations are:

$$
\begin{aligned}
\frac{dx}{dt} &= 2p \\
\frac{dy}{dt} &= 2q \\
\frac{dp}{dt} &= E_x(x,y) \\
\frac{dq}{dt} &= E_y(x,y).
\end{aligned}
\tag{IV.2.3}
$$

Since $E(x,y)$ vanishes precisely to second order at $(0,0)$, we can rewrite the characteristic equations in some neighborhood of the origin as:

$$
\begin{aligned}
\frac{dx}{dt} &= 2p \\
\frac{dy}{dt} &= 2q \\
\frac{dp}{dt} &= 2\alpha x + \beta y + o(|x| + |y|) \\
\frac{dq}{dt} &= \beta x + 2\gamma y + o(|x| + |y|).
\end{aligned}
\tag{IV.2.4}
$$

One can view x, y, p and q as coordinates in $\Re^4$. So, let ξ denote the four-tuple (x,y,p,q) and let $F(x,y,p,q) = p^2 + q^2 - E(x,y)$. We will say that $\xi \in F$ if x, y, p and q satisfy the equation $F = 0$. Using this notation we introduce:

Definition: Let $p^2 + q^2 = E(x,y)$ be a constrained eikonal equation and let ξ denote the four-tuple (x,y,p,q). The characteristic equations can be written as:

$$
\frac{d\xi}{dt} = A\xi + G(\xi) \tag{IV.2.5}
$$

where A is the four by four matrix:

$$A = \begin{pmatrix} 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \\ 2\alpha & \beta & 0 & 0 \\ \beta & 2\gamma & 0 & 0 \end{pmatrix} \tag{IV.2.6}$$

and where G has the following properites:

$$\begin{aligned} &1) \quad G(\xi) \in C^2 \\ &2) \quad G(0) = 0 \\ &3) \quad \frac{\partial G}{\partial \xi}(0) = 0 \end{aligned} \tag{IV.2.7}$$

Every solution $\xi = \xi(t)$ to (IV.2.5) is called an *orbit*. The equation:

$$\frac{d\xi}{dt} = A\xi \tag{IV.2.8}$$

is called the *linearized* characteristic equation.

To describe orbits in some neighborhood of an isolated singular point the following concept is useful:

Definition: Let

$$\frac{d\xi}{dt} = A\xi + G(\xi) \tag{IV.2.9}$$

be an ordinary differential equation as in (IV.2.5). An orbit $\xi = \xi(t)$ is *quasiradial* if:

$$\lim_{t \to \pm\infty} \xi(t) = 0. \tag{IV.2.10}$$

Equivalently we can express this definition as:

$$\lim_{t \to +\infty} (x(t), y(t)) = 0 \qquad \lim_{t \to +\infty} (p(t), q(t)) = 0 \tag{IV.2.11}$$

$$\lim_{t \to -\infty} (x(t), y(t)) = 0 \qquad \lim_{t \to -\infty} (p(t), q(t)) = 0. \tag{IV.2.12}$$

In other words, if the characteristic curves are quasiradial, they are quasiradial in the image plane and in gradient space. Some quasiradial orbits are depicted in figure 11.

The stable manifold theorem [ABMA80, p.527] and [HAR64, p.242] defines conditions under which the solutions to the linearized characteristic equations (IV.2.8) have

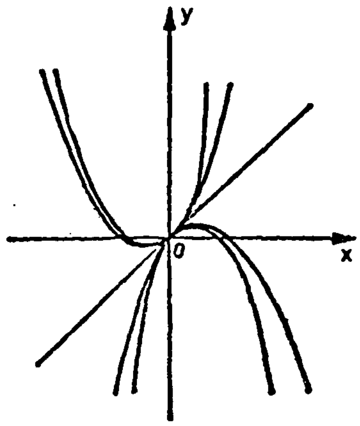


Figure 11. Quasiradial orbits

the same topological structure as the solutions to the characteristic equations (IV.2.5). One of the crucial constraints is that all eigenvalues of A have non-vanishing real part. A precise definition of a manifold can be found in [STO69, p.203]. Here it suffices to observe that a surface defined by a C^k function $z = z(x, y)$ is a two-dimensional C^k manifold. Before discussing the stable manifold theorem we state a theorem concerning two ordinary differential equations in two unknowns in some neighborhood of an isolated critical point:

Theorem: (Node Theorem [HAR64, p.213]) Let

$$\frac{dx}{dt} = a_{11}x + a_{12}y + f_1(x, y) \qquad \text{(IV.2.13)}$$
$$\frac{dy}{dt} = a_{21}x + a_{22}y + f_2(x, y)$$

be a system of two ordinary differential equations where $a_{i,j}$ for $i, j = 1, 2$ are constants such that $a_{11}a_{22} - a_{12}a_{21} \neq 0$ and f_i for $i = 1, 2$ are real continuous functions for which the following conditions hold in some neighborhood of the origin:

$$f_1 = o(|x| + |y|) \qquad f_2 = o(|x| + |y|) \qquad \text{(IV.2.14)}$$

Thus $(x, y) = (0, 0)$ is an isolated critical point. If both eigenvalues of the linearized equation of (IV.2.13) are real and have the same sign, then all orbits are quasiradial and the critical point is called a *node*. In particular if both eigenvalues are negative, then each orbit tends to the origin as $t \to +\infty$ and if both eigenvalues are positive, then each orbits tends to the origin as $t \to -\infty$.

The previous theorem can be viewed as a special case of the following theorem:

Theorem (Stable Manifold Theorem): Let

$$\frac{d\xi}{dt} = A\xi + G(\xi) \tag{IV.2.15}$$

be an ordinary differential equation where A is a constant matrix which has d_1 eigenvalues with negative real part and d_2 eigenvalues with positive real part, where d_1 and d_2 are both positive, and where $G(\xi)$ satisfies the following conditions:

$$\begin{aligned} &1) \quad G(\xi) \in C^2 \\ &2) \quad G(0) = 0 \\ &3) \quad \frac{\partial G}{\partial \xi}(0) = 0. \end{aligned} \tag{IV.2.16}$$

Then there exist an $\epsilon > 0$ such that (IV.2.15) has solutions $\xi = \xi(t) \neq 0$ satisfying:

$$||\xi(t)||e^{\epsilon t} \to 0 \quad \text{as} \quad t \to \infty \tag{IV.2.17}$$

and has solutions $\xi = \xi(t) \neq 0$ satisfying:

$$||\xi(t)||e^{\epsilon t} \to 0 \quad \text{as} \quad t \to -\infty \tag{IV.2.18}$$

Furthermore, for sufficiently small $\epsilon > 0$, the point $\xi = 0$ and the set of points $\xi(t)$ which satisfy (IV.2.17) sweep out a unique C^2 manifold of dimension d1, the *stable* manifold. Similarly, for sufficiently small $\epsilon > 0$, the point $\xi = 0$ and the set of points $\xi(t)$ which satisfy (IV.2.18) sweep out a unique C^2 manifold of dimension d2, the *unstable* manifold.

Thus the curves which sweep out the stable (unstable) manifold are quasiradial.

IV.3. Integral Surfaces near a Singular Point

We will now prove the main results of this chapter.

Lemma: Let $p^2 + q^2 = E(x, y)$ be a constrained eikonal equation. If a locally convex solution exists in some neighborhood of the singular point, then it is swept out by quasiradial characteristic curves.

Proof: Suppose a locally convex solution $z = z(x, y)$ exists. As $z = z(x, y)$ is assumed to be C^2, we can write p and q in some neighborhood of the singular point as:

$$\begin{aligned} p &= a_{11}x + a_{12}y + o(|x| + |y|) \\ q &= a_{12}x + a_{22}y + o(|x| + |y|) \end{aligned} \tag{IV.3.1}$$

where a_{ij} for $i, j = 1, 2$ are constants. As the origin is a singular point, p and q have no constant terms. Note that the Gaussian curvature K of $z = z(x, y)$ at the origin is defined by:

$$K = a_{11}a_{22} - a_{12}^2. \tag{IV.3.2}$$

Recall (section III.2.4) that a surface is locally convex at a point if its Gaussian curvature is positive there. Substituting the expressions (IV.3.1) into the first two characteristic equations (IV.2.3) of a constrained eikonal equation gives:

$$\begin{aligned} \frac{dx}{dt} &= 2(a_{11}x + a_{12}y) + o(|x| + |y|) \\ \frac{dy}{dt} &= 2(a_{12}x + a_{22}y) + o(|x| + |y|). \end{aligned} \tag{IV.3.3}$$

Using the node theorem we deduce that the characteristic curves in the x-y plane are quasiradial if and only if both eigenvalues of the linearized equations have the same sign. A simple calculation shows that this is the case only when $K > 0$. Similarly, we can show that the characteristic curves in gradient space are quasiradial if and only if $K > 0$ which in turn is true only if a_{11} and a_{22} have the same sign. Assuming that $K > 0$, the sign of the eigenvalues is the same as the sign of a_{11}. ∎

In the next section we will actually compute the coefficients a_{ij} for $i, j = 1, 2$ such that $K > 0$.

Finally we are ready to state and prove the main theorem of this chapter:

Theorem: Let $p^2 + q^2 = E(x, y)$ be a constrained eikonal equation. Then there exists a unique locally convex solution in some neighborhood of the singular point.

Proof: It follows from the previous lemma that if a locally convex solution to a constrained eikonal equation exists, it is swept out by quasiradial characteristic curves. So we have to show that such a solution exists and is unique. This is achieved by showing that the unstable manifold is the locally convex solution. To this end we investigate the linearized characteristic equations (IV.2.8) of a constrained eikonal equation. An easy calculation shows that the matrix A has two positive, real eigenvalues and two negative, real eigenvalues. Thus we can apply the stable manifold theorem, which states that there exist exactly two C^2 manifolds which are swept out by quasiradial characteristic curves. A locally convex solution therefore exists. From the node theorem we get that the solution $z = z(x, y)$ satisfying the positivity condition is the unstable manifold, whereas the stable manifold is the surface defined by $\tilde{z} = -z(x, y)$. Hence the locally convex solution is unique. ∎

IV.4. Formal Power Series Solution

In the previous section we proved that there exists a unique, locally convex solution to a constrained eikonal equation in some neighborhood of the singular point, but did not show how to actually compute such a solution. We will do so in this section for the case where $E(x,y)$ is analytic. In particular, we shall construct a *formal power series solution*. In the case where the eikonal equation is not analytic, such a solution tells us only about the behavior of an integral surface of the equation.

Definition: A *formal power series* is an expression of the form:

$$f = z_0 + \sum_{j=1}^{\infty} z_j \tag{IV.4.1}$$

where z_0 denotes a constant and for each j, z_j are homogeneous polynomials of degree j. We write the first order partial derivatives of f as:

$$\frac{\partial f}{\partial x} = \sum_{j=1}^{\infty} p_j \tag{IV.4.2}$$

$$\frac{\partial f}{\partial y} = \sum_{j=1}^{\infty} q_j$$

where for each j, p_j and q_j are homogeneous polynomials of degree j. We will say that $p^2 + q^2 = E(x,y)$ is a *super constrained eikonal equation* if $E(x,y)$ is C^∞. Then f is a *formal power series solution* to a super constrained eikonal equation if:

$$(\frac{\partial f}{\partial x})^2 + (\frac{\partial f}{\partial y})^2 = \alpha x^2 + \beta xy + \gamma y^2 + \sum_{j=1}^{\infty} e_j \tag{IV.4.3}$$

where for each j, e_j is a homogeneous polynomial of degree j.

Suppose for a moment that we have computed a formal power series solution to a constrained eikonal equation. A theorem due to Borel [HAR, p.261] states that there exists a C^∞ function $z(x,y)$ which has a given power series as its Taylor series expansion. Unfortunately f does not determine $z(x,y)$ uniquely since two different functions can have the same power series expansion. For example the functions $g(x,y)$ and $\tilde{g}(x,y)$ defined by (IV.4.4) have the same power series expansion at the origin:

$$g(x,y) = \sum_{j=1}^{\infty} x^j + \sum_{j=1}^{\infty} y^j \tag{IV.4.4}$$

$$\tilde{g}(x,y) = g(x,y) + e^{-\frac{1}{x^2}} + e^{-\frac{1}{y^2}}.$$

Note that the first order partial derivatives of $g(x,y)$ and $\tilde{g}(x,y)$ are the same only at the origin. Thus they cannot both be solutions to a given eikonal equation. As a formal power series solution satisfies the eikonal equation at the origin but not necessarily at any other point, it is not necessarily a solution to a given constrained eikonal equation. Yet the formal power series solution does tell us, qualitatively, about the behavior of a solution to an eikonal equation. Suppose a solution $\tilde{z}$ to the eikonal equation exists in some neighborhood of the origin. Then z and $\tilde{z}$ are tangent to at least second order at the origin.

> **Lemma**: Let $p^2 + q^2 = E(x,y)$ be a constrained eikonal equation where $E(x,y)$ is analytic. Then its formal power series solution is the solution to the equation.

Proof: A version of the stable manifold theorem proves that if $E(x,y)$ is analytic, then the stable (unstable) manifold is analytic [COLE55, p.330]. The lemma follows. ∎

> **Lemma**: Let $p^2 + q^2 = E(x,y)$ be a super constrained eikonal equation. Then there exists a unique, locally convex formal power series solution to this equation in some neighborhood of the singular point.

Proof: (outline) Equating the appropriate terms in (IV.4.3) we obtain

- an equation for the quadratic terms and
- a recurrence relation for each of the higher order terms.

First (section IV.4.1) we shall prove that there is a unique solution to the equation for the quadratic terms if we impose the constraints that the formal power series solution be positive and convex. The next step (section IV.4.2) is to determine the higher order terms which is done by inductively solving the recurrence relation. If the quadratic terms have been determined such that the formal power series solution is convex, each step of this induction can be carried out uniquely. The first of the following two equations determines the quadratic terms and the second is the recurrence relation from which the higher order terms can be calculated:

$$p_1^2 + q_1^2 = \alpha x^2 + \beta xy + \gamma y^2 \tag{IV.4.5}$$

$$2p_1 p_k + 2q_1 q_k = g_{k+1} \qquad \text{for} \quad k > 1 \tag{IV.4.6}$$

where for each k, g_k denotes a homogeneous polynomial of degree k. Each g_k is easily computed using the power series expansion of $E(x,y)$ as we will show in section IV.4.2. ∎

IV.4.1. Quadratic Terms

In this section we determine p_1 and q_1 for given values of α, β and γ in (IV.4.5). Note that these terms define the Gaussian curvature of a solution to a constrained eikonal equation. Since p_1 and q_1 are linear homogeneous polynomials they can be written as:

$$\begin{aligned} p_1 &= a_{11}x + a_{12}y \\ q_1 &= a_{21}x + a_{22}y \end{aligned} \tag{IV.4.7}$$

where the coefficients a_{ij} for $i, j = 1, 2$ are constants. As p_1 and q_1 satisfy equation (IV.4.5), the coefficients a_{ij} for $i, j = 1, 2$ are constrained by:

$$(a_{11}x + a_{12}y)^2 + (a_{21}x + a_{22}y)^2 = \alpha x^2 + \beta xy + \gamma y^2. \tag{IV.4.8}$$

Furthermore p_1 and q_1 are the linear terms of the first order partial derivatives of a smooth function. Thus the so-called compatibility condition has to hold:

$$\frac{\partial p}{\partial y} = \frac{\partial q}{\partial x} \tag{IV.4.9}$$

which constrains the coefficients a_{12} and a_{21}:

$$a_{12} = a_{21}. \tag{IV.4.10}$$

Equating appropriate terms in (IV.4.8) and using the compatibility condition we derive three equations for a_{11}, a_{12} and a_{22}:

$$\begin{aligned} a_{11}^2 + a_{12}^2 &= \alpha \\ 2a_{12}(a_{11} + a_{22}) &= \beta \\ a_{12}^2 + a_{22}^2 &= \gamma. \end{aligned} \tag{IV.4.11}$$

This system of equations only has a solution if both α and γ are not negative, which was assumed.

As we wish to show that there exists only one positive convex formal power series solution $z(x, y)$ to an eikonal equation, we want the coefficients a_{ij} for $i, j = 1, 2$ to satisfy the previous equations and K to be positive:

$$K = a_{11}a_{22} - a_{12}^2. \tag{IV.4.12}$$

Thus the Gaussian curvature of z at the origin is determined only by p_1 and q_1. We will say that a solution for the coefficients a_{ij} for $i, j = 1, 2$ is convex when K is positive. Note that for K to be positive it is necessary that a_{11} and a_{22} have the same sign.

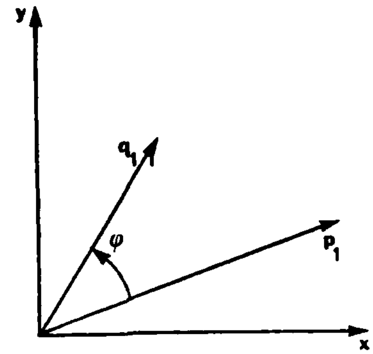


Figure 12. Vectors p_1 and q_1

We can view p_1 and q_1 as vectors in the x-y plane. Then the vector product of the vectors p_1 and q_1 is:

$$p_1 \times q_1 = a_{11}a_{22} - a_{12}^2. \tag{IV.4.13}$$

Recall that the vector product of two vectors can be written also as:

$$p_1 \times q_1 = |p_1||q_1| \sin \varphi \tag{IV.4.14}$$

where φ denotes the angle between the vectors p_1 and q_1 as shown in figure 12. Thus for p_1 and q_1 to define a convex surface, the angle between them has to be less than 180 degrees. We show now that the equations (IV.4.11) constrain this angle. The inner product of p_1 and q_1 is:

$$p_1 \cdot q_1 = a_{12}(a_{11} + a_{22}). \tag{IV.4.15}$$

Equivalently, we can write the inner product of p_1 and q_1 as:

$$p_1 \cdot q_1 = |p_1||q_1| \cos \varphi \tag{IV.4.16}$$

where φ again denotes the angle between the vectors p_1 and q_1. The following equations also hold:

$$|p_1| = |\sqrt{a_{11}^2 + a_{12}^2}| \tag{IV.4.17}$$
$$|q_1| = |\sqrt{a_{12}^2 + a_{22}^2}|.$$

So, we express the cosine of the angle φ between p_1 and q_1 as:

$$\cos \varphi = \frac{\beta}{2|\sqrt{\alpha\gamma}|} \tag{IV.4.18}$$

and there are two different angles φ_1 and $\varphi_2 = 2\pi - \varphi_1$ which satisfy this equation. The next equation defines $\sin\varphi$ in terms of a_{12}, α and γ:

$$\sin\varphi = \frac{(\pm\sqrt{\alpha - a_{12}^2})(\pm\sqrt{\gamma - a_{12}^2}) - a_{12}^2}{|\sqrt{\alpha\gamma}|}. \tag{IV.4.19}$$

Inasmuch as we are interested solely in convex solutions (i.e., $\varphi < 180°$), this equation has to be solved only for the case where the product of the square roots is positive, or equivalently, when a_{11} and a_{22} have the same sign. Combining the last two equations yields:

$$\frac{\sqrt{\alpha - a_{12}^2}\sqrt{\gamma - a_{12}^2} - a_{12}^2}{|\sqrt{\alpha\gamma}|} = \sqrt{1 - \frac{\beta^2}{4\alpha\gamma}} \tag{IV.4.20}$$

Since the the coefficients α, β and γ define a positive form, the previous equation is well defined and can be rewritten as:

$$4(\sqrt{4\alpha\gamma - \beta^2} + \alpha + \gamma)a_{12}^2 = \beta^2. \tag{IV.4.21}$$

It is possible to determine a_{12} in terms of α, β and γ from this equation as long as $\alpha \neq \gamma$ and $\beta \neq 0$. The case where $\alpha = \gamma$ and $\beta = 0$ will be investigated later.

Case 1) $\alpha \neq \gamma$ and $\beta \neq 0$

Let us denote the two possible solutions for a_{12} obtained from (IV.2.21) by $\bar{a}_{12}$ and $-\bar{a}_{12}$. Without loss of generality we assume that $\beta > 0$. The results for $\beta < 0$ are analogous. The two solutions for a_{11}, a_{12} and a_{22} are:

$$a_{11} = \sqrt{\alpha - \bar{a}_{12}^2} \qquad a_{12} = \bar{a}_{12} \qquad a_{22} = \sqrt{\gamma - \bar{a}_{12}^2} \tag{IV.4.22}$$

$$a_{11} = -\sqrt{\alpha - \bar{a}_{12}^2} \qquad a_{12} = -\bar{a}_{12} \qquad a_{22} = -\sqrt{\gamma - \bar{a}_{12}^2}. \tag{IV.4.23}$$

The coefficients defined by (IV.4.22) determine the unique positive convex solution.

Case 2) $\alpha = \gamma$ and $\beta = 0$

In this case equations (IV.4.11) can be written as:

$$a_{11}^2 + a_{12}^2 = \alpha \tag{IV.4.24}$$

$$2a_{12}(a_{11} + a_{22}) = 0 \tag{IV.4.25}$$

$$a_{12}^2 + a_{22}^2 = \alpha. \tag{IV.4.26}$$

Subtracting (IV.4.26) from (IV.4.24) we get:

$$a_{11}^2 - a_{22}^2 = 0 \tag{IV.4.27}$$

and deduce that:

$$a_{11} = \pm a_{22}. \tag{IV.4.28}$$

When $a_{11} = -a_{22}$, K is negative. So we only have to investigate the case where $a_{11} = a_{22}$. It follows from (IV.4.25) that $a_{12} = 0$, thus we can express the coefficients a_{ij} for $i, j = 1, 2$ in terms of α as:

$$\begin{aligned} a_{11} &= \pm\sqrt{\alpha} \\ a_{12} &= 0 \\ a_{22} &= \pm\sqrt{\alpha}. \end{aligned} \tag{IV.4.29}$$

Both solutions for the coefficients a_{ij} for $i, j = 1, 2$ are convex, but only one of them is positive. Thus we have shown that in the case where $\alpha = \gamma$ and $\beta = 0$ a unique positive convex solution exists. ∎

Actually we can also show the following theorem:

> **Theorem**: Let $p^2 + q^2 = E(x, y)$ be a constrained eikonal equation. Then at the singular point, the Gaussian curvature of each integral surface has the same absolute value and is determined by the (limited) Taylor series expansion of $E(x, y)$ at that singular point.

Proof: Recall that the curvature at the origin, denoted by K, is:

$$K = a_{11}a_{22} - a_{12}^2. \tag{IV.4.30}$$

Using equations (IV.4.11) we derive an expression for this curvature in terms of α, β, γ and a_{12}:

$$\begin{aligned} (a_{11} + a_{22})^2 &= \frac{\beta^2}{4a_{12}^2} \\ (a_{11} + a_{22})^2 &= \alpha - a_{12}^2 + \gamma - a_{12}^2 + 2a_{11}a_{22} \\ (a_{11} + a_{22})^2 &= \alpha + \gamma + 2K. \end{aligned} \tag{IV.4.31}$$

The desired expression for K is:

$$K = \frac{1}{2}\left[\frac{\beta^2}{4a_{12}^2} - (\alpha + \gamma)\right]. \tag{IV.4.32}$$

From equations (IV.4.18) and (IV.4.19) we get the two solutions for a_{12}^2 in terms of α, β and γ:

$$a_{12}^2 = \frac{\beta^2}{4[\sqrt{4\alpha\gamma - \beta^2} + (\alpha + \gamma)]} \tag{IV.4.33}$$
$$a_{12}^2 = \frac{\beta^2}{4[\sqrt{4\alpha\gamma - \beta^2} + (\alpha + \gamma)]}.$$

Substituting these two expressions for a_{12}^2 into *equation* (IV.4.32) *gives:*

$$|K| = \frac{1}{2}|\sqrt{4\alpha\gamma - \beta^2}|. \tag{IV.4.34}$$

Thus the absolute value of the curvature at the singular point of all integral surfaces of a constrained eikonal equation can be directly determined from the image intensities defined by $E(x, y)$. ∎

IV.4.2. Higher Order Terms

In this section we calculate the higher order terms in a formal power series solution to a constrained eikonal equation using the solutions for p_1 and q_1, and show that this can be done in a unique fashion if p_1 and q_1 determine a locally convex solution. First we derive *the recurrence relation* which has to be solved to determine the higher order terms. Suppose that p_1 and q_1 are known. For p_2 and q_2 the following equation must hold:

$$2p_1p_2 + 2q_1q_2 = l_3. \tag{IV.4.35}$$

For p_3 and q_3 the following equation holds:

$$2p_1p_3 + 2q_1q_3 + p_2^2 + q_2^2 = l_4. \tag{IV.4.36}$$

Now, by assumption p_2 and q_2 are homogeneous polynomials of degree two. We can therefore write the previous equation as:

$$2p_1p_3 + 2q_1q_3 = g_4 \tag{IV.4.37}$$

where g_4 can be determined from l_4, p_2 and q_2. Thus to determine, for each k, p_k and q_k we have to solve the following recurrence relation:

$$p_1p_{k-1} + q_1q_{k-1} = g_k \tag{IV.4.38}$$

where g_k is a known homogeneous polynomial of degree k.

We now introduce a new coordinate system denoted by X and Y:

$$X = p_1 \qquad Y = q_1. \tag{IV.4.39}$$

We can do so, as p_1 and q_1 are linearly independent vectors. (An easy calculation shows that if p_1 and q_1 are linearly dependent they define a surface which has zero curvature at the origin. In other words the coefficients α, β and γ would not define a positive form.) Thus for each $k > 2$, we can write (IV.4.38) as:

$$XP(X,Y) + YQ(X,Y) = F(X,Y) \tag{IV.4.40}$$

where $F(X,Y)$ is a homogeneous polynomial of degree k and $P(X,Y)$ and $Q(X,Y)$ are homogeneous polynomials of degree $k-1$. This follows immediately from the definition of X and Y. This last equation always has a solution, i.e.,:

$$\begin{aligned} \bar{P}(X,Y) &= \frac{F_X(X,Y)}{k} \\ \bar{Q}(X,Y) &= \frac{F_Y(X,Y)}{k} \end{aligned} \tag{IV.4.41}$$

since, using these two functions in (IV.4.40), we obtain:

$$XF_X(X,Y) + YF_Y(X,Y) = kF(X,Y) \tag{IV.4.42}$$

which is Euler's homogenity relation [BRSE75, p.246] for a homogeneous polynomial of degree k.

Now we show that any solution for $P(X,Y)$ and $Q(X,Y)$ can be written as:

$$\begin{aligned} P(X,Y) &= \bar{P}(X,Y) + C(X,Y)Y \\ Q(X,Y) &= \bar{Q}(X,Y) - C(X,Y)X \end{aligned} \tag{IV.4.43}$$

where $C(X,Y)$ is a homogeneous polynomial of degree $k-2$. Suppose that $P^1(X,Y)$, $Q^1(X,Y)$ and $P^2(X,Y), Q^2(X,Y)$ satisfy (IV.4.40). Then $\tilde{P}(X,Y) = P^1 - P^2$ and $\tilde{Q}(X,Y) = Q^1 - Q^2$ have to satisfy the following equation:

$$X\tilde{P}(X,Y) + Y\tilde{Q}(X,Y) = 0. \tag{IV.4.44}$$

We are looking for two functions $\tilde{P}(X,Y)$ and $\tilde{Q}(X,Y)$ which satisfy the last equation and are homogeneous polynomials of degree $k-1$. Furthermore $\tilde{P}(X,Y)$ must vanish at $Y = 0$ whereas $\tilde{Q}(X,Y)$ must vanish at $X = 0$. Thus $\tilde{P}(X,Y)$ and $\tilde{Q}(X,Y)$ must be of the form:

$$\begin{aligned} \tilde{P}(X,Y) &= \quad C(X,Y)Y \\ \tilde{Q}(X,Y) &= -C(X,Y)X \end{aligned} \tag{IV.4.45}$$

which proves our assertion (IV.4.43).

Hence we can write the functions $P(X,Y)$ and $Q(X,Y)$, which satisfy equation (IV.4.40), as:

$$\begin{aligned} P(X,Y) &= \frac{F_X(X,Y)}{k} + C(X,Y)Y \\ Q(X,Y) &= \frac{F_Y(X,Y)}{k} - C(X,Y)X. \end{aligned} \tag{IV.4.46}$$

Yet we want to find some $P(X,Y)$ and $Q(X,Y)$ which not only satisfy (IV.4.40) but also the compatibility condition:

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}. \tag{IV.4.47}$$

We will show that if p_1 and q_1 define a convex solution, only one homogeneous polynomial $C(X,Y)$ of degree $k-2$ exists such that P and Q satisfy this condition. Using (IV.4.46), the partial derivatives of P and Q with respect to x and y are:

$$\begin{aligned} \frac{\partial P}{\partial y} &= \frac{1}{k}(F_{XX}\frac{\partial X}{\partial y} + F_{XY}\frac{\partial Y}{\partial y}) + \frac{\partial Y}{\partial y}C(X,Y) + Y(\frac{\partial C}{\partial X}\frac{\partial X}{\partial y} + \frac{\partial C}{\partial Y}\frac{\partial Y}{\partial y}) \\ \frac{\partial Q}{\partial x} &= \frac{1}{k}(F_{YX}\frac{\partial X}{\partial x} + F_{YY}\frac{\partial Y}{\partial x}) - \frac{\partial X}{\partial x}C(X,Y) - X(\frac{\partial C}{\partial X}\frac{\partial X}{\partial x} + \frac{\partial C}{\partial Y}\frac{\partial Y}{\partial x}). \end{aligned} \tag{IV.4.48}$$

Recall the definitions of X and Y in terms of x and y:

$$\begin{aligned} X &= a_{11}x + a_{12}y \\ Y &= a_{12}x + a_{22}y. \end{aligned} \tag{IV.4.49}$$

So the partial derivatives of X and Y with respect to x and y are:

$$\begin{aligned} \frac{\partial X}{\partial x} &= a_{11} \qquad & \frac{\partial X}{\partial y} &= a_{12} \\ \frac{\partial Y}{\partial x} &= a_{12} \qquad & \frac{\partial Y}{\partial y} &= a_{22}. \end{aligned} \tag{IV.4.50}$$

As $F(X,Y)$ is a homogeneous polynomial, $F_{XY} = F_{YX}$. Using equations (IV.4.48) and (IV.4.50) we can express the compatibility condition (IV.4.47) as:

$$\begin{aligned} &\frac{1}{k}[F_{XX}a_{12} - F_{YY}a_{12} + F_{XY}(a_{22} - a_{11})] = \\ &- C(X,Y)(a_{11} + a_{22}) - C_X(Xa_{11} + Ya_{12}) - C_Y(Xa_{12} + Ya_{22}). \end{aligned} \tag{IV.4.51}$$

We wish to determine when there is only one polynomial $C(X,Y)$ of degree $k-2$ which satisfies the previous equation. In so doing, we will compute the coefficients in

the polynomial $C(X,Y)$. Then let

$$\begin{aligned}
F(X,Y) &= \sum_{i=0}^{k} f_{i,k-i} X^i Y^{k-i} \\
F_X(X,Y) &= \sum_{i=1}^{k} f_{i,k-i}\, i X^{i-1} Y^{k-i} \\
F_Y(X,Y) &= \sum_{i=0}^{k-1} f_{i,k-i}(k-i) X^i Y^{k-i-1} \\
F_{XX}(X,Y) &= \sum_{i=2}^{k} f_{i,k-i}\, i(i-1) X^{i-2} Y^{k-i} \\
F_{YY}(X,Y) &= \sum_{i=0}^{k-2} f_{i,k-i}(k-i)(k-i-1) X^i Y^{k-i-2} \\
F_{XY}(X,Y) &= \sum_{i=1}^{k-1} f_{i,k-i}\, i(k-i) X^{i-1} Y^{k-i-1}.
\end{aligned} \tag{IV.4.52}$$

Similarly, letting

$$\begin{aligned}
C(X,Y) &= \sum_{j=0}^{k-2} c_{j,k-j-2} X^j Y^{k-j-2} \\
C_X(X,Y) &= \sum_{j=1}^{k-2} c_{j,k-j-2}\, j X^{j-1} Y^{k-j-2} \\
C_Y(X,Y) &= \sum_{j=0}^{k-3} c_{j,k-j-2}(k-j-2) X^j Y^{k-j-3}
\end{aligned} \tag{IV.4.53}$$

gives:

$$\begin{aligned}
XC_X(X,Y) &= \sum_{j=1}^{k-2} c_{j,k-j-2}\, j X^j Y^{k-j-2} \\
YC_X(X,Y) &= \sum_{j=1}^{k-2} c_{j,k-j-2}\, j X^{j-1} Y^{k-j-1} \\
XC_Y(X,Y) &= \sum_{j=0}^{k-3} c_{j,k-j-2}(k-j-2) X^{j+1} Y^{k-j-3} \\
YC_Y(X,Y) &= \sum_{j=0}^{k-3} c_{j,k-j-2}(k-j-2) X^j Y^{k-j-2}.
\end{aligned} \tag{IV.4.54}$$

Using the last three systems of equations in (IV.4.51) we can rewrite the compatibility condition as:

$$
\begin{aligned}
&\frac{1}{k}[\sum_{i=2}^{k} f_{i,k-i}i(i-1)X^{i-2}Y^{k-i}a_{12} - \\
&\sum_{i=0}^{k-2} f_{i,k-i}(k-i)(k-i-1)X^{i}Y^{k-i-2}a_{12} + \\
&\sum_{i=1}^{k-1} f_{i,k-i}i(k-i)X^{i-1}Y^{k-i-1}(a_{22}-a_{11})] = \qquad \text{(IV.4.55)}\\
&-\sum_{j=0}^{k-2} c_{j,k-j-2}X^{j}Y^{k-j-2}(a_{11}+a_{22}) - \\
&\sum_{j=1}^{k-2} c_{j,k-j-2}jX^{j}Y^{k-j-2}a_{11} - \\
&\sum_{j=1}^{k-2} c_{j,k-j-2}jX^{j-1}Y^{k-j-1}a_{12} - \\
&\sum_{j=0}^{k-3} c_{j,k-j-2}(k-j-2)X^{j+1}Y^{k-j-3}a_{12} - \\
&\sum_{j=0}^{k-3} c_{j,k-j-2}(k-j-2)X^{j}Y^{k-j-2}a_{22}.
\end{aligned}
$$

Equating the coefficients of the powers $X^{\mu}Y^{k-\mu}$ for $\mu = 0, ..., k-2$ we obtain the following equations:

For $\mu = 0$:

$$
\begin{aligned}
&\frac{1}{k}[f_{2,k-2}2a_{12} - f_{0,k}k(k-1)a_{12} + f_{1,k-1}(k-1)(a_{22}-a_{11})] = \qquad \text{(IV.4.56)}\\
&- c_{0,k-2}(a_{11}+a_{22}) - c_{1,k-3}a_{12} - c_{0,k-2}(k-2)a_{22}
\end{aligned}
$$

For $0 < \mu < k-2$:

$$
\begin{aligned}
&\frac{1}{k}[f_{\mu+2,k-\mu-2}(\mu+2)(\mu+1)a_{12} - f_{\mu,k-\mu}(k-\mu)(k-\mu-1)a_{12} + \\
&f_{\mu+1,k-\mu-1}(k-\mu-1)(k-\mu)(a_{22}-a_{11})] = \qquad \text{(IV.4.57)}\\
&- c_{\mu,k-\mu-2}(a_{11}+a_{22}) - c_{\mu,k-\mu-2}\mu a_{11} - \\
&- c_{\mu+1,k-\mu-3}(\mu+1)a_{12} - c_{\mu-1,k-\mu-1}(k-\mu-1)a_{12} - \\
&c_{\mu,k-\mu-2}(k-\mu-2)a_{22}
\end{aligned}
$$

For $\mu = k-2$:

$$\frac{1}{k}[f_{k,0}k(k-1)a_{12} - f_{k-2,2}2a_{12} + f_{k-1,1}(k-1)(a_{22}-a_{11})] = \\ -c_{k-2,0}(a_{11}+a_{22}) - c_{k-2,0}(k-2)a_{11} - c_{k-3,1}a_{12}. \tag{IV.4.58}$$

Denote by F_μ, for $\mu = 0, ..., k-2$, the left hand side of each appropriate equation, multiplied by -1. Then we can rewrite (IV.4.56), (IV.4.57) and (IV.4.58) as:

$$\begin{aligned} F_0 =& c_{0,k-2}[a_{11} + a_{22}(k-1)] + c_{1,k-3}a_{12} \\ F_\mu =& c_{\mu-1,k-\mu-1}(k-\mu-1)a_{12} + \\ & c_{\mu,k-\mu-2}[a_{11}(\mu+1) + a_{22}(k-\mu-1)] + \\ & c_{\mu+1,k-\mu-3}(\mu+1)a_{12} \\ F_{k-2} =& c_{k-3,1}a_{12} + c_{k-2,0}[a_{11}(k-1) + a_{22}]. \end{aligned} \tag{IV.4.59}$$

Now, for each fixed $k > 2$ we need to determine the coefficients $c_{i,k-j-2}$ from the previous system of $k-1$ linear equations. This system can be written in matrix form as:

$$A_{k-1}c = F \tag{IV.4.60}$$

and has a unique solution if the matrix A_{k-1} is nonsingular (i.e., its determinant is not equal to zero). Note that A_{k-1} is a tridiagonal band matrix, so its determinant can be computed using the recurrence relation (IV.4.61) which we now derive.

Let B_n be an $n \times n$ matrix, let B_{n-1} denote the submatrix of B_n obtained from B_n by eliminating its top row and leftmost column and let B_{n-2} be similarly obtained from B_{n-1}. Then B_n can be written as:

$$B_n = \begin{pmatrix} a & b & 0 \\ c & B_{n-1} & \\ 0 & & \end{pmatrix}. \tag{IV.4.61}$$

The determinant of B_n may be computed in terms of the determinants of its submatrices:

$$\det B_n = a \det B_{n-1} - bc \det B_{n-2}. \tag{IV.4.62}$$

It suffices to prove by induction that A_{k-1} is nonsingular if $a_{11}a_{22} - a_{12}^2 > 0$. Without loss of generality (see section IV.4.1) we can assume that $a_{11} > 0$ and $a_{22} > 0$. The μth row of A_{k-1} is:

$$(k-\mu-1)a_{12} \qquad a_{11}(\mu+1) + a_{22}(k-\mu-1) \qquad a_{12}(\mu+1). \tag{IV.4.63}$$

Set $j = k-\mu-2$ and denote by Δ_j the determinant of the submatrix of A_{k-1} consisting of its $j+1$ bottom rows and $j+1$ rightmost columns. We prove that A_{k-1} is nonsingular by induction on j.

For the basis case, we have to show that $\Delta_0 \neq 0$ and $\Delta_1 \neq 0$:

$$\Delta_0 = a_{11}(k-1) + a_{22} \tag{IV.4.64}$$
$$\Delta_1 = [a_{11}(k-2) + 2a_{22}][a_{11}(k-1) + a_{22}] - a_{12}^2(k-2).$$

By assumption Δ_0 is positive. Rewrite Δ_1 as:

$$\Delta_1 = a_{11}^2(k-2)(k-1) + 2a_{11}a_{22}(k-1) + 2a_{22}^2 + (a_{11}a_{22} - a_{12}^2)(k-2). \tag{IV.4.65}$$

Since we have assumed that $a_{11}a_{22} - a_{12}^2 > 0$, it follows that Δ_1 is always positive.

Assume now that Δ_{j-2} and Δ_{j-1}, for $j > 1$, are positive. We show that Δ_j is then also positive which completes the proof. Using (IV.4.62) we get:

$$\Delta_j = [a_{11}(k-j-1) + a_{22}(j+1)]\Delta_{j-1} - a_{12}^2(k-j-1)j\Delta_{j-2}. \tag{IV.4.66}$$

where Δ_{j-1} is defined by:

$$\Delta_{j-1} = [a_{11}(k-j) + a_{22}j]\Delta_{j-2} - a_{12}^2(k-j)(j-1)\Delta_{j-3} \tag{IV.4.67}$$

and where we define Δ_{-1} to be 1. Using (IV.4.67) in (IV.4.66) gives:

$$\begin{aligned}\Delta_j =& a_{11}(k-j-1)\{[a_{11}(k-j) + a_{22}j]\Delta_{j-2} - a_{12}^2(k-j)(j-1)\Delta_{j-3}\} + \\ & a_{22}(j+1)\Delta_{j-1} - a_{12}^2(k-j-1)j\Delta_{j-2} \end{aligned} \tag{IV.4.68}$$
$$\begin{aligned}\Delta_j =& (a_{11}a_{22} - a_{12}^2)(k-j-1)j\Delta_{j-2} + a_{22}(j+1)\Delta_{j-1} + \\ & a_{11}(k-j-1)(k-j)[a_{11}\Delta_{j-2} - a_{12}^2(j-1)\Delta_{j-3}]. \end{aligned} \tag{IV.4.69}$$

To verify that Δ_j is positive, one need only show, therefore, that:

$$a_{11}\Delta_{j-2} > a_{12}^2(j-1)\Delta_{j-3}. \tag{IV.4.70}$$

We proceed, once again, by induction. For $j = 3$ we wish to show:

$$a_{11}\Delta_1 > 2a_{12}^2\Delta_0. \tag{IV.4.71}$$

Using (IV.4.64) and (IV.4.65) in (IV.4.71) this can be equivalently written as:

$$\begin{aligned} a_{11}[a_{11}^2(k-2)(k-1) + 2a_{11}a_{22}(k-1) + 2a_{22}^2 + (a_{11}a_{22} - a_{12}^2)(k-2)] > \\ 2a_{12}^2[a_{11}(k-1) + a_{22}]. \end{aligned} \tag{IV.2.72}$$

To see that the previous equation holds it suffices to observe that:

$$\begin{aligned} 2a_{11}^2a_{22}(k-1) &> 2a_{12}^2a_{11}(k-1) \\ 2a_{11}a_{22}^2 &> 2a_{12}^2a_{22}. \end{aligned} \tag{IV.4.73}$$

This solves the basis case. Assume, by induction, that the next inequality holds:

$$a_{11}\Delta_{j-2} > a_{12}^2(j-1)\Delta_{j-3}. \tag{IV.4.74}$$

We need to show that:

$$a_{11}\Delta_{j-1} > a_{12}^2 j\Delta_{j-2}. \tag{IV.4.75}$$

Using (IV.4.67) in this last inequality gives:

$$a_{11}^2(k-j)\Delta_{j-2} + a_{11}a_{22}j\Delta_{j-2} - a_{11}a_{12}^2(k-j)(j-1)\Delta_{j-3} > a_{12}^2 j\Delta_{j-2}. \tag{IV.4.76}$$

This holds if:

$$a_{11}^2(k-j)\Delta_{j-2} > a_{11}a_{12}^2(k-j)(j-1)\Delta_{j-3} \tag{IV.2.77}$$

which is true by the induction hypothesis (IV.4.74).

Hence there exists a unique locally convex formal power series solution to a constrained eikonal equation which can be easily computed. ∎

Chapter V

B-Silhouettes

V.1. Overview of Basic Concepts

As discussed, our goal is to find necessary constraints such that an image can be interpreted in a unique way when its image irradiance equation is known. It was shown in chapter III that, in general, there can be an infinite number of different surfaces which satisfy the same image irradiance equation; in other words, the image of each of these surfaces is the same for a fixed imaging configuration. Recall that in sections III.2.2 and III.3 we discussed how to detect a b-silhouette in an image. It remains now to investigate whether the existence of such b-silhouettes can be used to interpret an image. In this chapter we will identify three constraints upon an image irradiance equation, one upon the reflectance map, one upon the b-silhouette and one upon the function $E(x, y)$. If these constraints hold for some image irradiance equation, only one surface defined by a C^2 function which satisfies the equation exists.

What kind of information about an integral surface can one deduce from a singular image irradiance equation? Let $R(p, q) = E(x, y)$ be a fixed, singular image irradiance equation whose nondegenerate b-silhouette is defined by $w(x, y) = 0$. As previously discussed (section III.3), the surface normal to the bounding contour of an integral surface is parallel to $(\pm w_x, \pm w_y, 0)$.

V.2. Uniqueness Theorem

In this section we obtain constraints which assure that if an image irradiance

equation has a C^2 integral surface, it is unique. So let

$$R(p,q) = E(x,y) \tag{V.2.1}$$

be a singular image irradiance equation. Consider the following constraints upon this equation:

C1) $R(p,q) = p^2 + q^2$.

C2) The b-silhouette defined by $w(x,y) = 0$ is a closed, smooth curve in the x-y plane. Furthermore, the points (x,y) at which the image irradiance equation is defined, lie in the region bounded by this b-silhouette.

C3) The function $E(x,y)$ has exactly one stationary point (x_0, y_0) and satisfies the following conditions in some neighborhood of (x_0, y_0): $E(x_0, y_0) = 0$, $E(x,y) > 0$ for $(x,y) \neq (x_0, y_0)$ and $E(x,y)$ vanishes precisely to second order at (x_0, y_0).

Uniqueness Theorem: Let $R(p,q) = E(x,y)$ be an image irradiance equation for which constraints C1, C2 and C3 hold and suppose a C^2 integral surface defined by $z = z(x,y)$ of this equation exists. Then the only other solution to the equation is $\tilde{z} = -z(x,y)$.

Proof: Let $R(p,q) = E(x,y)$ be a fixed image irradiance equation for which constraints C1, C2 and C3 hold. First note that the point $P = (x,y,p,q) = (x_0, y_0, 0, 0)$ is an isolated singular point of the image irradiance equation (see also section IV.2). There are then two observations which allow us to prove the theorem. First, as the b-silhouette is a closed curve, an integral surface of the equation has to be compact. Second, from constraints C1 and C3 we can deduce that such a surface is convex at the singular point, which allows us to apply results of the previous chapter.

Suppose $z = z(x,y)$ defines a C^2 integral surface of an image irradiance equation as defined in the uniqueness theorem. Then from C2 we may infer that z defines a compact surface (section III.2.4). Note also that $z(x,y)$ is defined for every point (x,y) which lies within or on the b-silhouette and therefore has a bounding contour. Thus there exists a point $\tilde{P}$ at which z has an extremum. In particular the tangent plane at $\tilde{P}$ is parallel to the x-y plane.

From condition C3 we can deduce that there exists a point (x_0, y_0, z_0) on z such that the plane tangent to z at this point is parallel to the x-y plane, i.e., $p(x_0, y_0) = 0$ and $q(x_0, y_0) = 0$. By assumption this is the only singular point of the image irradiance equation and therefore the only point on z where the tangent plane is parallel to the x-y plane. Furthermore, as the image irradiance equation is singular, the point (x_0, y_0) lies in the interior of the region of the x-y plane bounded by the b-silhouette. Hence

$\tilde{P} = (x_0, y_0, z_0)$. By the assumption on $E(x, y)$, z is either convex at $\tilde{P}$ or has a saddle point. Since $\tilde{P}$ is the point where the surface has maximal (minimal) height, z must be convex there.

In chapter IV we proved that if an image irradiance equation satisfies C1 and C3, there exists only one positive and exactly one negative convex solution denoted by z_p and $z_n = -z_p$ respectively. Thus there are exactly two integral surfaces $z = z(x, y)$ and $\tilde{z} = -z(x, y)$. ∎

By using transformation methods, we can enlarge the class of singular image irradiance equations for which the uniqueness theorem holds. Let

$$f(Ap^2 + 2Bpq + Cq^2 + 2Dp + 2Eq) = E(x, y) \qquad \text{(V.2.2)}$$

be a singular image irradiance equation where f is a bijection and A, B, C, D and E are real constants such that $\delta > 0$ and $\Delta S < 0$ where δ, Δ and S are defined in the following equations:

$$\begin{aligned} \delta &= AC - B^2 \\ \Delta &= \begin{vmatrix} A & B & D \\ B & C & E \\ D & E & 0 \end{vmatrix} \\ S &= A + C. \end{aligned} \qquad \text{(V.2.3)}$$

The constraints upon the constants A, B, C, D and E in equation (V.2.2) assure that the curves $R(p, q) = c$, for any constant c, are closed. Let the b-silhouette of the equation be a closed and smooth curve. Then (V.2.2) can be transformed into an image irradiance equation of the form (V.2.1) for which C2, holds as is shown in appendix II. If, after the transformation, the function $E(x, y)$ satisfies C3, then the uniqueness theorem holds for (V.2.2).

The next corollary follows directly from the uniqueness theorem in this chapter. We will abbreviate $\sqrt{x^2 + y^2}$ by r.

> **Corollary**: Let $p^2 + q^2 = E(r)$ be an image irradiance equation where $E(r)$ satisfies constraints C2 and C3. Suppose a C^2 integral surface of this image irradiance equation exists. Then it is *rotationally symmetric* and can be obtained by integrating $E(r)$. In this case the b-silhouette is a circle.

Proof: First we write the eikonal equation in polar coordinates:

$$z_r^2 + \frac{1}{r^2} z_\vartheta^2 = E(r). \qquad \text{(V.2.4)}$$

Let $\tilde{z} = \tilde{z}(r)$ define the rotationally symmetric integral surface of the above eikonal equation. Thus $\tilde{z}_\vartheta(r) = 0$ and we can compute both rotationally symmetric solutions by integrating $\pm\sqrt{E(r)}$. It follows from the uniqueness theorem that the image irradiance equation has only rotationally symmetric integral surfaces. ∎

Note that the above corollary does not hold if the image does not contain a b-silhouette. In section III.2.5 we showed that the integral surfaces of a continuous rotationally symmetric eikonal equation are not themselves necessarily rotationally symmetric. The uniqueness result for the special image irradiance equation (III.3.4) has been independently obtained by [DSY80].

V.3. Counterexamples

In the previous section we discussed sufficient constraints under which the solution to a singular image irradiance equation is unique. Are these constraints necessary? Although we are not able to answer this question completely, we now shed some light upon it. In particular we try to find the class of image irradiance equations for which most likely there is no set of constraints that assure the existence of only one global solution.

Image irradiance equations satisfying the constraints of the uniqueness theorem have closed iso-brightness curves, i.e., the curves $R(p,q) = c$ are closed. So let us examine singular image irradiance equations whose iso-brightness curves are not closed. Such an image irradiance equation is given by:

$$p + q = \frac{-(x+y)}{\sqrt{1-(x^2+y^2)}}. \tag{V.3.1}$$

While constraint C2 holds for (V.3.1), an image irradiance equation where the reflectance map is $R(p,q) = p + q$ never has a singular point. The general solution to (V.3.1) is:

$$z(x,y) = \sqrt{1-(x^2+y^2)} + w(y-x) \tag{V.3.2}$$

where w is any C^1 function. Figures 13 through 16 illustrate some solutions to equation (IV.3.1).

Another example of an image irradiance equation whose iso-brightness curves are not closed is given by:

$$pq = \frac{xy}{1-(x^2+y^2)}. \tag{V.3.3}$$

This equation satisfies constraint C2, i.e., its b-silhouette is a closed and smooth curve. Furthermore $(0,0)$ is the singular point of (V.3.3) and $E(x,y)$ vanishes precisely to second order there although $E(x,y)$ is not positive in the neighborhood of $(0,0)$. One

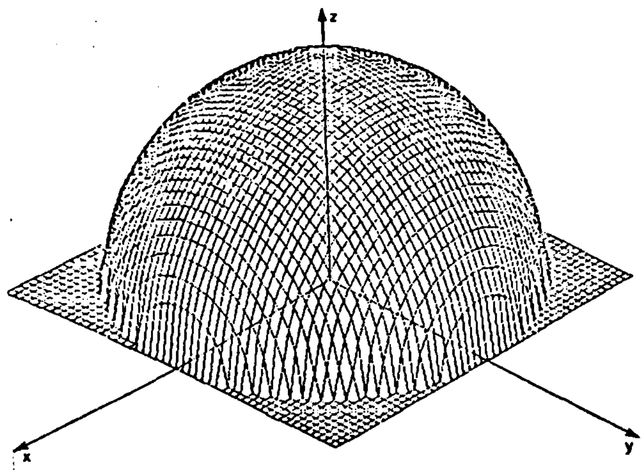


Figure 13. $z(x,y) = \sqrt{1-(x^2+y^2)}$

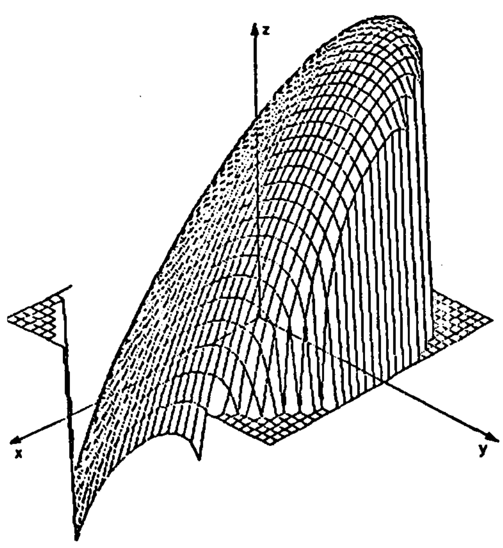


Figure 14. $z(x,y) = \sqrt{1-(x^2+y^2)} + (y-x)$

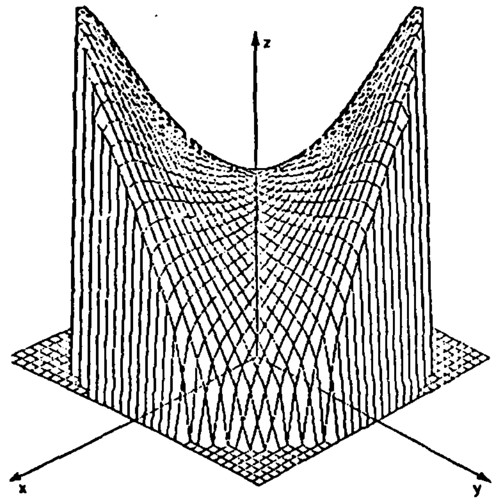


Figure 15. $z(x, y) = \sqrt{1 - (x^2 + y^2)} + (y - x)^2$

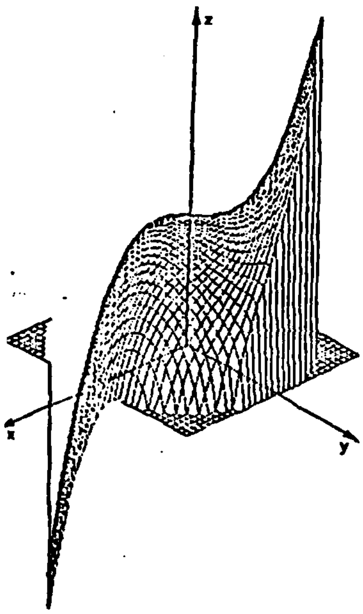


Figure 16. $z(x, y) = \sqrt{1 - (x^2 + y^2)} + (y - x)^3$

of the solutions to (V.3.3) is the sphere:

$$z(x,y) = \sqrt{1-(x^2+y^2)} \tag{V.3.4}$$

whereas another surface satisfying (V.3.3) is:

$$\begin{aligned} z(x,y) &= f(t) + x^2 - y^2 \quad \text{where} \\ t &= 1-(x^2+y^2) \qquad \text{and} \\ f(t) &= t\sqrt{\frac{1}{4t}+1} + \frac{1}{8}[\ln(\sqrt{\frac{1}{4t}+1}+1) - \ln(\sqrt{\frac{1}{4t}+1}-1)]. \end{aligned} \tag{V.3.5}$$

Recall that constraint C2 expressed the fact that the b-silhouette is a closed and smooth curve. We now demonstrate that if the b-silhouette does not obey C2, our uniqueness result does not hold. Previously, it was not expected that the uniqueness results for image irradiance equations containing closed b-silhouettes would be different from the results for those containing open b-silhouettes. An example of an equation for which C1 holds, but whose b-silhouette is not a closed curve is:

$$p^2 + q^2 = \frac{1}{4x} + 1. \tag{V.3.6}$$

Equation (V.3.6) does not have a singular point. A solution to this equation is:

$$z(x,y) = \sqrt{x} + y \tag{V.3.7}$$

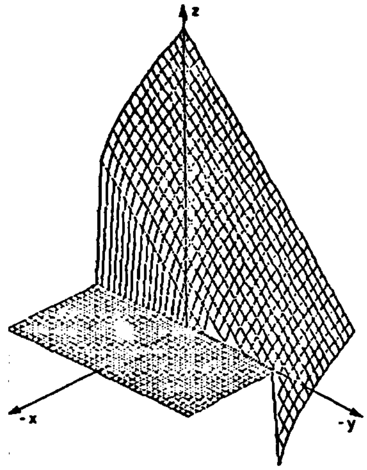


Figure 17. $z(x,y) = \sqrt{x} + y$

which is shown in figure 17. Two other solutions to (V.3.6) are:

$$z(x,y) = x\left(\sqrt{\frac{1}{4x}+1}\right) + \frac{1}{8}\left[\ln\left(\sqrt{\frac{1}{4x}+1}+1\right) - \ln\left(\sqrt{\frac{1}{4x}+1}-1\right)\right]. \qquad \text{(V.3.8)}$$

$$z(x,y) = \frac{\sqrt{x(1-8x)}}{2} - \frac{1}{4\sqrt{2}}\operatorname{atan}\sqrt{\frac{1}{8x}-1} + \sqrt{3}y$$

A more complicated counterexample is:

$$p^2 + q^2 = \frac{x^2+y^2}{1-xy}. \qquad \text{(V.3.9)}$$

This equation clearly satisfies C1 but its b-silhouette is not a closed curve. The function $E(x,y)$ vanishes precisely to second order at the singular point of (V.3.9) and is positive in the neighborhood of the origin. Two different solutions to (V.3.9) are given by equation (V.3.10) (shown in figure 18) and (V.3.11) (shown in figure 19).

$$z(x,y) = 2\sqrt{1-xy} \qquad \text{(V.3.10)}$$

$$z(x,y) = (1-xy)\sqrt{\frac{1}{1-xy}-1} - \operatorname{atan}\sqrt{\frac{1}{1-xy}-1} + \frac{x^2-y^2}{2} \qquad \text{(V.3.11)}$$

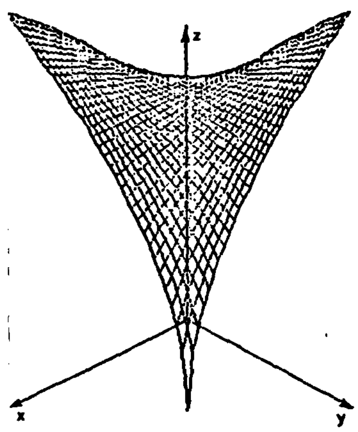


Figure 18. $z(x,y) = 2\sqrt{1 - xy}$

Note that the surface defined by (V.3.11) is C^1 at the origin (although not C^2) and is only defined in two quadrants, whereas the image irradiance equation is defined in all four quadrants. The surface defined by equation (V.3.10) is not convex at the origin. Only when the b-silhouette is a closed curve can we deduce that a surface is convex at the singular point, an observation which allows us to prove the uniqueness theorem. For the following eikonal equation we give two solutions which are both C^2 at the origin but only one of which is convex:

$$p^2 + q^2 = \frac{x^2 + y^2}{1 - x^2 y^2}. \tag{V.3.12}$$

Equation (V.3.12) has a singular point and $E(x,y)$ satisfies constraint C3. Two solutions to this equation are:

$$\begin{aligned} z(x,y) &= \arcsin(xy) \\ z(x,y) &= \sqrt{1 - x^2 y^2} + \frac{x^2 - y^2}{2}. \end{aligned} \tag{V.3.13}$$

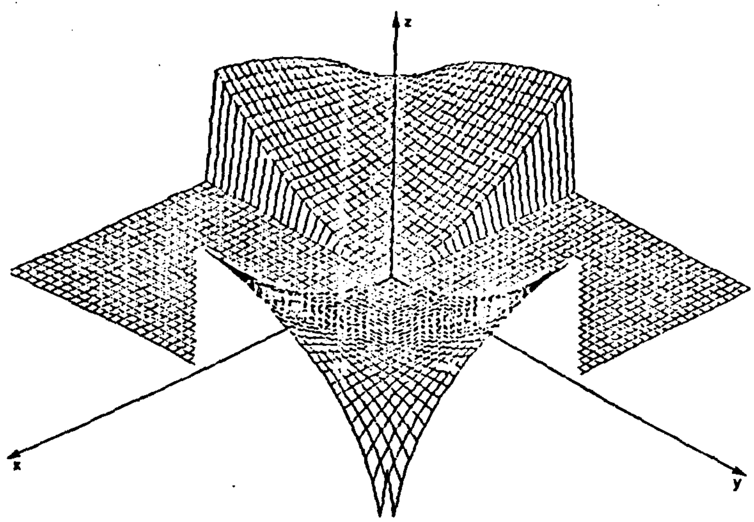


Figure 19. $z(x,y) = (1 - xy)\sqrt{\frac{1}{1-xy} - 1} - \operatorname{atan}\sqrt{\frac{1}{1-xy} - 1} + \frac{x^2 - y^2}{2}$

Chapter VI

Numerical Methods

VI.1. General Remarks

In this chapter we discuss some numerical methods proposed to solve the shape from shading problem. The first shape from shading algorithm was implemented by Horn [HO70], who proposed to solve the characteristic equations (III.2.2) using standard numerical methods for solving ordinary differential equations. Horn observes that the surface gradient at a singular point P (section III.2.2 and chapter IV) is uniquely defined by the image irradiance equation. By assuming that a surface is locally convex in some neighborhood of P and then estimating the curvature of this surface, he is able to calculate an initial curve (sections A.6 and A.7). Horn notes that using this heuristic, only one surface is calculated. However, it is not guaranteed that there exists a solution to any given image irradiance equation which is locally convex in some neighborhood of P and so the surface determined using Horn's algorithm may not be an integral surface. Furthermore, the algorithm computes at most one of the [illegible] integral surfaces of an image irradiance equation. We proved in chapter IV [illegible] case of an eikonal equation where some technical conditions are imposed on $E(x, y)$ a unique locally convex solution exists in some neighborhood of a singular point. In this situation then, Horn's algorithm computes the unique convex solution. (Note that using our results it is not necessary to estimate the curvature at a singular point in order to uniquely compute the surface.) Unfortunately, the algorithm is slow, numerically unstable, and relies on the presence of singular points.

Strat [STR79] developed an iterative shape from shading algorithm which we discuss in detail in section VI.2. One of the major shortcomings of this algorithm is that an initial curve lying on a surface is required as an input.

Brooks [BRO79] suggests a *Waltz-like* [WA75] constraint propagation scheme to determine the shape of a surface from its shading. Like Strat, Brooks assumes that a surface is C^2 and that an initial curve which is embedded in the surface is known. In addition, he imposes an upper limit upon the curvature of a surface that can be determined. This constraint (which stems from experiments with human visual systems) is needed for the algorithm to converge.

The three methods mentioned above can only solve the shape from shading problem under the assumption that a surface is smooth and does not have a bounding contour. If an image contains a b-silhouette it can be analyzed by using the algorithm due to Ikeuchi and Horn [IKHO81] which is explained in section VI.3.

VI.2. Strat's Algorithm

Strat [STR79] proposes an iterative algorithm to solve an image irradiance equation under the assumption that an integral surface is defined by a C^2 function $z = z(x, y)$. He assumes that image irradiance is measured at discrete points on the image plane and that the reflectance map is given. In this scheme, a square grid is imposed on the x-y plane. We denote a grid point by the tuple (i, j), the image irradiance measured at (i, j) by $E_{i,j}$, the first order partial derivatives of $z = z(x, y)$ (which defines an integral surface) at a point (i, j) by $p_{i,j}$ and $q_{i,j}$, and the reflectance map evaluated at $(p_{i,j}, q_{i,j})$ by $R_{i,j}$. The objective of the algorithm is to calculate $p_{i,j}$ and $q_{i,j}$ at every point (i, j). Strat does not address the problem of how to compute the function which defines the surface, i.e., $z_{i,j}$, from the values of $p_{i,j}$ and $q_{i,j}$.

So let $R(p, q) = E(x, y)$ be a given image irradiance equation. We now show how to obtain a function e (called the *error function*) in the variables $p_{i,j}$ and $q_{i,j}$ which is zero when the image irradiance equation is satisfied at every point and $p_{i,j}$ and $q_{i,j}$ are the first order partial derivatives of $z = z(x, y)$. Strat's algorithm computes iteratively the values for $p_{i,j}$ and $q_{i,j}$ which minimize the function e. This function is the weighted sum (over all (i, j)) of two functions $\epsilon^r_{i,j}$ and $\epsilon^s_{i,j}$ whose derivations we now explain. The function $\epsilon^r_{i,j}$, defined by (VI.2.1), is zero when $p_{i,j}$ and $q_{i,j}$ satisfy the image irradiance equation:

$$\epsilon^r_{i,j} = (E_{i,j} - R_{i,j})^2. \qquad \text{(VI.2.1)}$$

The values for $p_{i,j}$ and $q_{i,j}$ which minimize $\epsilon^r_{i,j}$ cannot, however, be chosen independently; they are the first order partial derivatives of a C^2 function $z = z(x, y)$. Consequently, equation (VI.2.2) holds for p and q [DIR72]:

$$p_y = q_x. \qquad \text{(VI.2.2)}$$

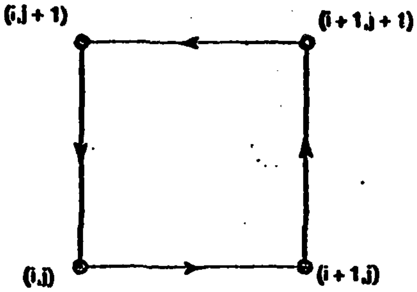


Figure 20. Template

Strat assumes that the image irradiance equation is solved over a simply connected region in the x-y plane. Hence (VI.2.2) can also be expressed in integral form [DIR72]:

$$\oint pdx + qdy = 0. \qquad \text{(VI.2.3)}$$

We now show how a discrete approximation of equation (VI.2.3) is used to derive $\epsilon^s_{i,j}$. Strat suggests the loop around four adjacent points on the grid as a path over which the integral in (VI.2.3) should be taken. Such a path is depicted in figure 20. A discrete, first order approximation to equation (VI.2.3) is then given by:

$$p_{i,j} + p_{i+1,j} + q_{i+1,j} + q_{i+1,j+1} - p_{i+1,j+1} - p_{i,j+1} - q_{i,j+1} - q_{i,j} = 0. \quad \text{(VI.2.4)}$$

Thus the values for $p_{i,j}$ and $q_{i,j}$ which the algorithm calculates must not only satisfy the image irradiance equation, but also equation (VI.2.4). In general, a point (i,j) belongs to four templates. For each template, a discrete approximation of (VI.2.3) can be derived and $p_{i,j}$ and $q_{i,j}$ must satisfy each such approximation. So $\epsilon^s_{i,j}$ is defined as:

$$\begin{aligned}
\epsilon^s_{i,j} =&(p_{i+1,j} + q_{i+1,j} + q_{i+1,j+1} - p_{i+1,j+1} - \\
&p_{i,j+1} - q_{i,j+1} + p_{i,j} - q_{i,j})^2 + \\
&(p_{i,j-1} + p_{i+1,j-1} + q_{i+1,j-1} + q_{i+1,j} - \\
&p_{i+1,j} - q_{i,j-1} - p_{i,j} - q_{i,j})^2 + \\
&(p_{i-1,j-1} + p_{i,j-1} + q_{i,j-1} - p_{i-1,j} - \\
&q_{i-1,j} - q_{i-1,j-1} - p_{i,j} + q_{i,j})^2 + \\
&(p_{i-1,j} + q_{i,j+1} - p_{i,j+1} - p_{i-1,j+1} - \\
&q_{i-1,j+1} - q_{i-1,j} + p_{i,j} + q_{i,j})^2.
\end{aligned} \qquad \text{(VI.2.5)}$$

The total (global) error e is then defined as:

$$e = \sum_{i,j} (\epsilon^s_{i,j} + \rho \epsilon^r_{i,j}) \tag{VI.2.6}$$

where ρ is a constant scalar chosen appropriately to equate the dimensions and to weigh the magnitudes of the errors $\epsilon^s_{i,j}$ and $\epsilon^r_{i,j}$. At each iteration, Strat's algorithm minimizes e. Hence at every point (i, j) the following two equations must be solved for $p_{i,j}$ and $q_{i,j}$:

$$\frac{\partial e}{\partial p_{i,j}} = 0 \qquad \frac{\partial e}{\partial q_{i,j}} = 0. \tag{VI.2.7}$$

The Gauss Seidel method is used to solve equations (VI.2.7).

It remains to describe how the initial assignment for the values $p_{i,j}$ and $q_{i,j}$ is chosen. Strat assumes that the algorithm is applied to solve an image irradiance equation over a rectangular region in the x-y plane. For points (i, j) on the boundary of the rectangle, the values for $p_{i,j}$ and $q_{i,j}$ are given as input data and for interior points, $p_{i,j}$ and $q_{i,j}$ are set to zero. As mentioned in section VI.1, Strat's method is not very useful in practice, since it requires, as input, the initial values for $p_{i,j}$ and $q_{i,j}$ at the boundary of the domain in which the algorithm is applied. It is not shown whether the solution computed by the algorithm is independent of this initial choice of values for $p_{i,j}$ and $q_{i,j}$ at the interior points.

Experimental results appear to support the conjecture that Strat's algorithm converges, although a proof of convergence is not given. One of the shortcomings of Strat's algorithm is that its performance depends upon the order in which the grid points are scanned, i.e., the order in which the values for $p_{i,j}$ and $q_{i,j}$ are updated.

VI.3. Ikeuchi and Horn's Algorithm

Ikeuchi and Horn [IKHO81] designed and implemented an iterative algorithm which differs in various respects from the ones described in the previous sections.

- Their assumption about the surface whose shading is to be analyzed is weaker. They observe that for a surface to look smooth it suffices for the function defining it to be only C^1. Recall that under this assumption, an integral surface of an image irradiance equation can be built from characteristic strips (see chapter III and appendix I).

- They show that a b-silhouette can be used as initial data, which is not possible using Brooks' or Strat's algorithm.

One of the strongest features of the method is that it incorporates data obtained from a b-silhouette. Recall that the surface normal at a point on the bounding contour is parallel to the normal vector to the b-silhouette. Before proceeding to describe and analyze Ikeuchi and Horn's algorithm, we review their slightly different formulation of an image irradiance equation.

In chapter III it was shown that an image irradiance equation defines a relation between the image plane and gradient space. In the original formulation of this equation [HO75] the (p, q) coordinate system was chosen to specify gradient space. With this underlying coordinate system an image irradiance equation can be formulated as a FOPDE. On the other hand this system has the disadvantage that for points on the bounding contour p and/or q assume infinite value. In order to find a numerical solution to an image irradiance equation it would be useful to transform gradient space into a different space SP, where (u, v) is the underlying coordinate system, so that the components of a surface normal vector (which always have finite value), instead of p and q, can be computed. To achieve this we have to find a space SP and a mapping m between gradient space and SP, satisfying the following constraints:

- The mapping m between gradient space and a region of SP is one-to-one and onto.
- The mapping m maps the whole (unbounded) p-q plane into a bounded region in SP.

Recall (section III.2.5) that gradient space can be obtained by projection of the Gaussian hemisphere from its center onto a plane placed at its south pole. Ikeuchi and Horn observe that by choosing a different projection of the Gaussian hemisphere onto a plane, each point on this hemisphere gets mapped into a point other than infinity on the plane. Several projections are suggested, one of them being the stereographic projection. We can think of this projection in geometric terms as a projection onto a plane tangent to the hemisphere at the south pole. The center of projection is the north pole, not the center of the sphere. In this case the resulting coordinate system, denoted by (f, g), is defined in terms of p and q by:

$$f = \frac{2p(\sqrt{1+p^2+q^2}-1)}{p^2+q^2} \tag{VI.3.1}$$

$$g = \frac{2q(\sqrt{1+p^2+q^2}-1)}{p^2+q^2}. \tag{VI.3.2}$$

We can now write an image irradiance equation in the form:

$$R(f, g) = E(x, y). \tag{VI.3.3}$$

Next we describe Ikeuchi and Horn's method of solving the shape from shading problem. The basic concepts of this algorithm are similar to those described in section VI.2. The inputs to this algorithm are the measured image irradiance and the

reflectance map $R(f,g)$ and the desired outputs are the values $f_{i,j}$ and $g_{i,j}$. (The problem of how to determine $z_{i,j}$ from the values of $f_{i,j}$ and $g_{i,j}$ is not addressed by Ikeuchi and Horn.) There are two major differences between this and Strat's procedure. First, a different error function is used, and second the initial values for $f_{i,j}$ and $g_{i,j}$ are computed for points on the b-silhouette. (The case where there are no b-silhouettes in the image is discussed later.) For points not on the b-silhouette, $f_{i,j}$ and $g_{i,j}$ are initialized to zero just as in Strat's algorithm.

We now explain Ikeuchi and Horn's derivation of the error function. They exploit the assumption that the function $z = z(x,y)$ defining an integral surface is C^1. A function $F = F(x,y)$ is *continuous* at a point (x_0, y_0) if, for any $\epsilon > 0$, there exists a δ such that, for any (x,y) which lies in the circle defined by $(x-x_0)^2 + (y-y_0)^2 < \delta$, we have:

$$|F(x,y) - F(x_0,y_0)| < \epsilon. \tag{VI.3.4}$$

A discrete approximation of equation (VI.3.4) is given by:

$$|F_{i+1,j} + F_{i-1,j} + F_{i,j+1} + F_{i,j-1} - 4F_{i,j}| < 4\epsilon. \tag{VI.3.5}$$

The error function e is therefore defined to be:

$$e = \sum_{i,j}(e^s_{i,j} + \lambda e^r_{i,j}) \qquad \text{where} \tag{VI.3.6}$$

$$e^s_{i,j} = (E_{i,j} - R_{i,j})^2 \tag{VI.3.7}$$

$$\begin{aligned} e^r_{i,j} = &(f_{i+1,j} + f_{i-1,j} + f_{i,j+1} + f_{i,j-1} - 4f_{i,j})^2 + \\ &(g_{i+1,j} + g_{i-1,j} + g_{i,j+1} + g_{i,j-1} - 4g_{i,j})^2. \end{aligned} \tag{VI.3.8}$$

In equation (VI.3.6) λ denotes a scalar factor chosen appropriately to equate the dimensions and to weigh the magnitudes of the errors $e^s_{i,j}$ and $e^r_{i,j}$.

As in Strat's algorithm, the function e is minimized. Thus, at every iteration the next two equations are solved for $f_{i,j}$ and $g_{i,j}$:

$$\frac{\partial e}{\partial f_{i,j}} = 0 \tag{VI.3.9}$$

$$\frac{\partial e}{\partial g_{i,j}} = 0.$$

Once again the Gauss Seidel method is used to solve equations (VI.3.9). A lower bound on the number of iterations needed using this method is proportional to the mesh size but an upper bound is not known. However, numerical experiments indicate that indeed the algorithm converges.

As mentioned before, $f_{i,j}$ and $g_{i,j}$ are initialized to zero at every point except those on the b-silhouette. It is not shown that the algorithm is independent of these initial values.

In the case where there are no b-silhouettes in the image, heuristics have to be used to initialize the algorithm. Ikeuchi and Horn do not specify the number of points at which the exact values for $f_{i,j}$ and $g_{i,j}$ need to be known in order to guarantee that the algorithm will compute an answer.

Chapter VII

Conclusion

In this report, we have investigated the question of how much information concerning the shape of an object can be deduced from its shaded image. Even assuming that adequate data is available to derive an image irradiance equation is insufficient to solve the reconstruction problem uniquely; in general, for a fixed imaging configuration there are many surfaces which have the same shaded image. Thus our goal has been to identify constraints by which the reconstruction problem can be solved uniquely.

We first analyzed the continuous image irradiance equation. The information needed to restrict the solutions to such an equation to a single one, was previously known [COHI62b]: If a strip is specified which is not a characteristic strip and along which $\Delta \neq 0$ (III.2.3), then there exists a unique surface which contains this strip and whose image has a particular shading for a fixed imaging configuration. We were interested in whether edges could constitute an initial curve. In the case where the image irradiance equation is linear, it is not possible to distinguish among the multiple surfaces which could give rise to a known edge. For nonlinear image irradiance equations an edge constrains the possible surfaces to a small number. However, if there exists a surface which contains a vertex and edges emanating from it, it is unique.

Then we discussed how singular points of an eikonal equation constrain its possible solutions. In particular we proved that there exists a unique (up to translation in the z-direction) positive convex surface which satisfies an eikonal equation in some neighborhood of a singular point.

Finally we investigated images in which b-silhouettes can be detected. An image irradiance equation which describes the relationship between the gradient of a surface whose image contains a b-silhouette, and the shading of this surface, can be singular.

We showed that singular image irradiance equations may be solved using the same methods applied to find the integral surfaces of a continuous image irradiance equation.

However, our *ultimate ambition was to answer the following question:*

Is there a set of constraints which assure that if an image irradiance equation has a solution, it is unique?

We answered this question affirmatively in chapter V. It was shown there that if three constraints are known to hold, the information about the imaging situation and the surface as captured by an image irradiance equation, allow one to reconstruct the shape of the surface in a unique manner. Furthermore one can easily check whether an image irradiance equation satisfies these constraints. It is surprising that our uniqueness theorem holds only when the b-silhouette is a closed curve (constraint C2).

In order to evaluate the usefulness of our uniqueness theorem, we need to know which of the commonly arising image irradiance equations actually obey the above mentioned restrictions. In his paper on hill-shading [HO79], Horn discusses eighteen different reflectance maps which are applied to solve that problem. Constraint C1 holds for five of those reflectance maps. These equations are of the form:

$$R(p,q) = f(p^2 + q^2) \tag{VII.1.1}$$

where f is a bijection. A reflectance map of the form (VII.1.1) describes, for instance, the situation when the object is Lambertian and the light source and the viewer have the same position. Furthermore eikonal equations can be also used to automatically analyze images taken by a scanning electron microscope which seems to be one of the prime applications of our uniqueness theorem.

There are several issues not treated here which would be interesting to investigate further: mutual illumination, shadows and specularities, for example. Another open problem is to determine the class of functions $R = R(p,q)$ which are reflectance maps. Our only restriction upon these functions has been that they be C^1. If a smaller set can be found which is known to contain all reflectance maps, better methods could be developed to solve the shape from shading problem in practice. A further research problem is to find a good numerical shape from shading algorithm.

References

[ABMA80] Abraham, R. and J.E. Marsden, *Foundations of Mechanics,* The Benjamin/Cummings Publishing Company, (1980).

[ABRO67] Abraham, R. and J. Robbin, *Transversal Mappings and Flows,* W.A. Benjamin, Inc., (1967).

[AL42] Alexandroff, A., "Existence and Uniqueness of a Convex Surface with a Given Integral Curvature," *Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS* **XXXV**, 5 (1942).

[ARN78] Arnold, V.I., *Mathematical Methods of Classical Mechanics,* Springer Verlag, (1978).

[BRO79] Brooks, M.J., "Surface-Normals from Closed Path," *Proceedings of the Sixth International Joint Conference on Artificial Intelligence,* Vol. 1, Tokyo, (1979).

[BRSE75] Bronstein, I.N. and K.A. Semendjajew, *Taschenbuch der Mathematik,* BSB B.G. Teubner Verlagsgesellschaft, (1975).

[BRU79] Bruss, A.R., *Some Properties of Discontinuities in the Image Irradiance Equation,* A.I.Memo 517, Artificial Intelligence Laboratory, M.I.T., (1979).

[CAR76] Do Carmo, M.P., *Differential Geometry of Curves and Surfaces,* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, (1976).

[COHI62a] Courant, R. and D. Hilbert, *Methods of Mathematical Physics,* Vol. I, Interscience Publishers, (1962), 164–274.

[COHI62b] Courant, R. and D. Hilbert, *Methods of Mathematical Physics,* Vol. II, Interscience Publishers, (1962), 1–147.

[COLE55] Coddington, E.A. and N. Levinson, *Theory of Ordinary Differential Equations,* McGraw-Hill Book Company, Inc., (1955), 371–377.

[COP75] Copson, E.T., *Partial Differential Equations,* Cambridge University Press, (1975).

[DES72] Descovich, J., *Hőhere Mathematik,* Lecture Notes, Technische Universität Wien, (1972).

[DIR72] Dirschmid, H.-J., *Mathematik A3,* Lecture Notes, Technische Universität Wien, (1972).

[DSY80] Deift, P. and J. Sylvester, *Some Remarks on the Shape from Shading Problem in Computer Vision,* (1980), personal communication.

[EP62] Epstein, B., *Partial Differential Equations: An Introduction,* McGraw Hill Book Company Inc., (1962).

[GAR64] Garabedian, P.R., *Partial Differential Equations,* John Wiley & Sons, Inc., (1964).

[GOGU73] Golubitsky, M. and V. Guillemin, *Stable Mappings and Their Singularities,* Springer Verlag, (1973).

[HA28] Haar, A., "Über die Eindeutigkeit und Analyzität der Lősungen Partieller Differentialgleichungen," *Atti del congr. intern. dei mat.* 3,Bologna, (1928), 5–10.

[HAD64] Hadamard, J., *La Théorie des Equations aux Derivées Partielles,* Editions Scientifiques, Pekin, (1964).

[HAR64] Hartman, P., *Ordinary Differential Equations,* John Wiley & Sons, Inc., (1964).

[HICV52] Hilbert, D. and S. Cohn-Vossen, *Geometry and the Imagination,* Chelsea Publishing Company, New York, (1952).

[HO70] Horn, B.K.P., *Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from one View,* TR-79, Project MAC, M.I.T., (1970).

[HO75] Horn, B.K.P., "Obtaining Shape from Shading Information," *The Psychology of Computer Vision, P. H. Winston (ed.),* McGraw Hill Book Company Inc., (1975), 115–155.

[HO77] Horn, B.K.P., "Understanding Image Intensities," *Artificial Intelligence* 8, (1977), 201–231.

[HO79] Horn, B.K.P., "Hill-Shading and the Reflectance Map," *Proceedings: Image Understanding Workshop,* (April 1979).

[HOS79] Horn, B.K.P. and R.W. Sjoberg, "Calculating the Reflectance Map," *Applied Optics* 18, (1979), 1770–1779.

[IKHO81] Ikeuchi, K. and B.K.P. Horn, "Numerical Shape from Shading and Occluding Boundaries," *Artificial Intelligence* **15**, Special Issue on Computer Vision (1981).

[JOH78] John, F., *Partial Differential Equations,* Springer Verlag, Third Edition, (1978).

[LAWH60] Laponsky, A.B. and N.R. Whetten, "Dependence of Secondary Emission from MgO Single Crystals on Angle of Incidence," *Physics Review* **120**, 3 (1980).

[LER57] Leray, J., "Uniformisation de la Solution du Problème Linéaire Analytique de Cauchy près de la Variété qui Porte les Données de Cauchy," *Bull. Soc. Math. France* 85, (1957), 389–429.

[MA77] Marr, D., "Analysis of Occluding Contours," *Proceedings Royal Society London B* **197**, (1977), 441–475.

[MAC73] Mackworth, A.K., "Interpreting Pictures of Polyhedral Scenes," *Artificial Intelligence* 4, (1973), 121–137.

[MY48] Mŷskis, A., "The Uniqueness of the Solution of Cauchy's Problem," *Uspehi Matem. Nauk (N.S.)* **3**, 2(24) (1948), 3–46, (in russian); an english abstract can be found in *Mathematical Reviews*, 1949, 302–303.

[PLI54] Plis, A., "Characteristics of Nonlinear Partial Differential Equations," *Bulletin de L'Académie Polonaise des Sciences Cl.III.* **II**, 9 (1954).

[POG73] Pogorelov, A.V., *Extrinsic Geometry of Convex Spaces,* American Mathematical Society, Providence, Rhode Island, (1973).

[POG78] Pogorelov, A.V., *The Minkowski Multidimensional Problem,* V.H. Winston & Sons, Washington, D.C., (1978).

[RO71] Robbin, J.W., "Stable Manifolds of Semi-Hyperbolic Fixed Points," *Illinois Journal of Mathematics* **15**, (1971).

[SMI68a] Smirnow, W.I., *Lehrgang der Höheren Mathematik,* Vol. 2, VEB Deutscher Verlag der Wissenschaften, Berlin, (1968), 141–149.

[SMI68b] Smirnow, W.I., *Lehrgang der Höheren Mathematik,* Vol. 4, VEB Deutscher Verlag der Wissenschaften, Berlin, (1968), 259–325.

[STO69] Stoker, J.J., *Differential Geometry,* Wiley-Interscience, (1969).

[STR79] Strat, M.T., *A Numerical Method for Shape-From-Shading From a Single Image,* Working Paper 191, Artificial Intelligence Laboratory, M.I.T., (1979).

[WA75] Waltz, D., "Understanding Line Drawings of Scenes with Shadows," *The Psychology of Computer Vision, P. H. Winston (ed.),* McGraw Hill Book Company Inc., (1975), 19–91.

[WOOD78] Woodham, R.J., *Reflectance Map Techniques for Analyzing Surface Defects in Metal Casting,* TR-457, Artificial Intelligence Laboratory, M.I.T., (1978).

[WU66] Wunderlich, W., *Darstellende Geometrie I,* Bibliographisches Institut, Mannheim, Wien, Zürich, (1966).

Appendix I

Mathematical Details

This appendix contains a summary of results known about first order partial differential equations. A thorough treatment of this material can be found in [COHI62b, pp.62–153] and [SMI68b, pp.259–301].

A.1. Basics

For simplicity of exposition, we will discuss only first order partial differential equations involving a function z of two variables, x and y. The results can be generalized to functions of n variables in a straightforward fashion. Let p and q denote the first order partial derivatives of z with respect to x and y respectively. Then the relation:

$$F(x,y,z,p,q) = 0 \tag{A.1.1}$$

where F is a function of x, y, z, p and q, is called a first order partial differential equation (abbreviated in the following by FOPDE). The relation (A.1.1) is a *linear* FOPDE if it is linear in p and q with coefficients depending only on x and y and (A.1.1) is *quasi-linear* if it is linear in p and q with coefficients depending on x, y and z.

A function $z(x,y)$ is called a solution to (A.1.1) if in some region of the x-y plane the function and its derivatives identically satisfy the equation in x and y. Such a function is also called an *integral surface*.

The *general* solution to a FOPDE is a whole set of solutions, each of which satisfies the equation. Given a FOPDE, what kind of constraints can be imposed so that there is only one integral surface which satisfies both the constraints and the equation? Such constraints are for example boundary conditions or initial values. In section A.4 we will

state precisely what comprises a general solution and in sections A.6 and A.7 discuss what kind of constraints are necessary to pin down a particular solution.

Unless otherwise stated, we assume that F, z and all relevant derivatives are continuous.

A.2. The Quasi-Linear FOPDE

We consider quasi-linear FOPDE's first as their geometric interpretation is rather clear and so the relevant method for solving them can be explained and understood easily. In this case we rewrite the relation (A.1.1) as:

$$a(x,y,z)p + b(x,y,z)q = c(x,y,z). \tag{A.2.1}$$

Furthermore we assume that:

$$a^2 + b^2 \neq 0. \tag{A.2.2}$$

Suppose that the solutions to (A.2.1) are written implicitly as:

$$G(x,y,z) = 0. \tag{A.2.3}$$

Differentiating (denoted by subscripts) (A.2.3) with respect to x and y yields:

$$G_x + G_z z_x = 0 \qquad \text{and} \qquad G_y + G_z z_y = 0 \tag{A.2.4}$$

or equivalently:

$$p = -\frac{G_x}{G_z} \qquad \text{and} \qquad q = -\frac{G_y}{G_z}. \tag{A.2.5}$$

Using these equations in (A.2.1) we obtain:

$$a(x,y,z)G_x(x,y,z) + b(x,y,z)G_y(x,y,z) + c(x,y,z)G_z(x,y,z) = 0. \tag{A.2.6}$$

Note that, in general, (A.2.1) is a nonlinear FOPDE for the function $z(x,y)$ whereas (A.2.6) is a linear FOPDE for $G(x,y,z)$. We can interpret the coefficients a, b and c in (A.2.6) as the components of a vector field which is defined by $\xi = \xi(x,y,z) = [a(x,y,z), b(x,y,z), c(x,y,z)]$. Then we can rewrite (A.2.6) as:

$$(\xi, \nabla G) = 0 \tag{A.2.7}$$

where ∇G denotes the gradient of G and $(\,,\,)$ the inner product of two vectors. Equation (A.2.7) expresses the fact that ξ is perpendicular to ∇G. Since the vector ∇G is perpendicular to the surface defined by $G(x,y,z) = 0$ at each point (x,y,z), we deduce that ξ lies in the tangent plane of this integral surface at that point.

A *field line* is a curve whose tangent at every point has the same direction as the field vector there. An integral surface of (A.2.7) can be built up from field lines (called *characteristic curves* in this context) of the vector field ξ. To reiterate the previous statements: The tangent at each point of a characteristic curve has the same direction there as the vector ξ and therefore, by virtue of (A.2.7), is perpendicular to the surface normal to the integral surface $G(x,y,z)=0$. This does not mean that each quasi-linear FOPDE has a single solution. Such a FOPDE only constrains the possible orientations of the tangent planes at each point to a one-parameter manifold. As (A.2.1) is *linear* in p and q at every point of any integral surface, all feasible tangent planes intersect in a line which is called the *Monge axis.*

We describe now a method for finding characteristic curves. Such curves can be defined as functions of one parameter s: $x=x(s), y=y(s)$ and $z=z(s)$. The vector $[x(s),y(s),z(s)]$ is denoted by χ. Then $\frac{d\chi}{ds}=[\frac{dx(s)}{ds},\frac{dy(s)}{ds},\frac{dz(s)}{ds}]$ has the same direction as ξ and so the outer product of $\frac{d\chi}{ds}$ and ξ must equal zero:

$$\begin{aligned} b\frac{dz}{ds}-c\frac{dy}{ds}&=0\\ c\frac{dx}{ds}-a\frac{dz}{ds}&=0\\ a\frac{dy}{ds}-b\frac{dx}{ds}&=0. \end{aligned} \tag{A.2.8}$$

The relations (A.2.8) are normally written as:

$$dx{:}dy{:}dz=a{:}b{:}c. \tag{A.2.9}$$

The solutions to the equations (A.2.9) comprise a two-parameter family of curves in space (a family of characteristic curves), yet only a one-parameter subset of them generates the solutions to the FOPDE. To find this subset, an arbitrary function between the two free parameters obtained when solving (A.2.9) is introduced. Hence the general solution to (A.2.1) contains this arbitrary function. So each surface produced by a one-parameter family of characteristic curves is an integral surface. Conversely we now show that each integral surface is generated in such a fashion.

On each integral surface $z=z(x,y)$ the equations:

$$\frac{dx}{ds}=a(x,y,z) \qquad \frac{dy}{ds}=b(x,y,z) \tag{A.2.10}$$

define a one-parameter family of curves: $x=x(s), y=y(s), z=z(x(s),y(s))$. Note that on any curve in this family:

$$\frac{dz}{ds}=c(x,y,z) \tag{A.2.11}$$

as:

$$\frac{dz}{ds} = \frac{\partial z}{\partial x}\frac{dx}{ds} + \frac{\partial z}{\partial y}\frac{dy}{ds} = ap + bq = c. \tag{A.2.12}$$

Thus every integral surface is swept out by a one-parameter family of characteristic curves. The following example illustrates these ideas.

Example:

The following FOPDE is to be solved:

$$F(x, y, z, p, q) = xp + yq - z = 0. \tag{A.2.13}$$

As $a(x, y, z) = x$, $b(x, y, z) = y$ and $c(x, y, z) = 1$, the equations for the characteristic curves (A.2.9) are:

$$dx{:}dy{:}dz = x{:}y{:}1 \tag{A.2.14}$$

and have as their solution the two-parameter curves in space:

$$\begin{aligned} y &= C_1 x \\ z &= C_2 x \end{aligned} \tag{A.2.15}$$

where C_1 and C_2 are constants. Any integral surface can be built from the curves described by equations (A.2.15) and every such surface is a one-parameter manifold of these curves. Each of these manifolds is determined through coupling C_1 and C_2 by an arbitrary relation w:

$$\frac{z}{x} = C_2 = w(C_1) = w(\frac{y}{x}). \tag{A.2.16}$$

Thus the solution to the FOPDE is:

$$z = w(\frac{y}{x})x. \tag{A.2.17}$$

Writing this in parametric form gives:

$$y = C_1 x \qquad \text{and} \qquad z = w(C_1)x. \tag{A.2.18}$$

A.3. The General FOPDE

We can apply methods similar to those developed in the previous section to solve a general FOPDE:

$$F(x, y, z, p, q) = 0 \tag{A.3.1}$$

where we require that:

$$F_p^2 + F_q^2 \neq 0. \tag{A.3.2}$$

Our goal is to transform the problem of finding a solution to (A.3.1) into the problem of integrating a set of ordinary differential equations. Again, geometrical reasoning will help to find these equations.

Let P be a given point on an integral surface. Then the quantities p and q, which determine the direction of a tangent plane at P, are constrained by (A.3.1) to a one-parameter family of curves. (In other words: once the coordinates x, y and z of a point P are fixed, (A.3.1) is an equation for p and q. To write this equation in parametric form only one parameter is needed.) The envelope of the tangent planes is a conical surface which can have several sheets and is called the *Monge cone*. (A conical surface is produced by moving a straight line which is fixed at one point, along a curve.) "The considerations here refer merely to a suitable small range of tangent planes, e.g., a portion of a sheet of the cone where q can be expressed as a single-valued differentiable function of p" [COHI62b, p.75]. Each generator of this cone represents a possible direction of the tangent plane at P and is called a *characteristic direction*. Thus the integral surface has to fit into the field of Monge cones, i.e., has to always be tangent to them.

Recall now that in the quasi-linear case, a Monge cone degenerates to a Monge axis. To determine the solution in that case, we find the characteristic curves which at every point have as their tangent direction the direction of the Monge axis there. An integral surface is swept out by these curves. In the case of a general FOPDE, the same basic idea works. Again, we have to find the curves which at every point have as their tangent direction a characteristic direction. Let such a curve (called *focal curve*) be given by $x = x(s), y = y(s)$ and $z = z(s)$. Remember that a one-parameter family of such curves should sweep out an integral surface $z(x, y)$ of a given FOPDE; in other words, the functions $x(s), y(s), z(s), p(s)$ and $q(s)$ have to satisfy this FOPDE. The focal curves determine $x(s), y(s)$ and $z(s)$. Yet, (A.3.1) gives only one relationship between p and q and so in order to determine p and q, another equation is needed. This equation can be obtained by requiring that focal curves be *embedded* in an integral surface. (A focal curve is embedded if in some neighborhood of the projection of this curve onto the x-y plane, z is a single-valued, twice continuous differentiable function of x and y.) Focal curves which satisfy this last condition are called characteristic curves and a one-parameter family of characteristic curves sweeps out an integral surface.

The problem which actually has to be solved is that of finding an integral surface. So the solution to this proceeds in the opposite direction from that described in the preceding paragraphs. First, a set of equations, called the *characteristic equations* has to be found. The characteristic curves are a subset of the solutions of this set, from which an integral surface can be built up. In the following paragraphs the technical prerequisites needed to find an integral surface of (A.3.1) are developed.

As a first task we determine the equation of the Monge cone. Again, let P be a fixed point which has the coordinates (x, y, z). Then p and q, which satisfy (A.3.1),

are written as functions of a parameter u and all feasible tangent planes at (x, y, z) are expressed as:

$$(Z - z) = (X - x)p(u) + (Y - y)q(u). \tag{A.3.3}$$

The envelope of the planes described by the previous equation defines the Monge cone, a conical surface whose vertex is (x, y, z). The equation of the Monge cone is found by eliminating u from (A.3.3) and from (A.3.4) which is obtained by differentiating (A.3.3) with respect to u:

$$0 = \frac{(X - x)dp}{du} + \frac{(Y - y)dq}{du}. \tag{A.3.4}$$

Differentiating (A.3.1) with respect to the parameter u gives:

$$0 = F_p \frac{dp}{du} + F_q \frac{dq}{du}. \tag{A.3.5}$$

Assuming that neither $\frac{dp}{du}$ and $\frac{dq}{du}$ nor F_p and F_q vanish simultaneously, we derive the following equation from (A.3.3), (A.3.4) and (A.3.5):

$$\frac{X - x}{F_p} = \frac{Y - y}{F_q} = \frac{Z - z}{pF_p + qF_q}. \tag{A.3.6}$$

By substituting all possible values for p and q (i.e., all values for p and q which satisfy (A.3.1)) we obtain all generators of the Monge cone at the point (x, y, z). "Space curves having a characteristic direction at each point shall be called focal curves" [COHI62b, p.76]. Therefore focal curves have to satisfy the following differential equations:

$$\frac{dx}{ds} = F_p \quad . \quad \frac{dy}{ds} = F_q \qquad \frac{dz}{ds} = pF_p + qF_q. \tag{A.3.7}$$

To show that every integral surface z can be generated in such a fashion, let $z = z(x, y)$ be an integral surface on which p and q is also known. Then the equations:

$$\frac{dx}{ds} = F_p \qquad \frac{dy}{ds} = F_q \tag{A.3.8}$$

define a one-parameter family of curves. On these curves:

$$\frac{dz}{ds} = p\frac{dx}{ds} + q\frac{dy}{ds} \tag{A.3.9}$$

holds and using (A.3.8) in (A.3.9) we obtain the third equation of (A.3.7):

$$\frac{dz}{ds} = pF_p + qF_q. \tag{A.3.10}$$

The above equation is called the strip condition. "It states that the functions $x(s), y(s), z(s), p(s)$ and $q(s)$ not only define a space curve, but simultaneously a plane tangent to it at every point. A configuration consisting of a curve and a family of tangent planes to this curve is called a strip" [COHI62b, p.77].

Equation (A.3.9) states that the curves defined by (A.3.7) are focal curves. Now it is also required that a focal curve be embedded in an integral surface. Differentiating (A.3.1) with respect to x and y we obtain:

$$\begin{aligned} F_p p_x + F_q q_x + F_z p + F_x &= 0 \\ F_p p_y + F_q q_y + F_z q + F_y &= 0. \end{aligned} \tag{A.3.11}$$

Using (A.3.7) and the fact that $p_y = q_x$ leads to the following equations:

$$\begin{aligned} \frac{dp}{ds} &= p_x \frac{dx}{ds} + p_y \frac{dy}{ds} = p_x F_p + q_x F_q \\ \frac{dq}{ds} &= q_x \frac{dx}{ds} + q_y \frac{dy}{ds} = p_y F_p + q_y F_q. \end{aligned} \tag{A.3.12}$$

Finally, using (A.3.11) in (A.3.12) yields:

$$\begin{aligned} \frac{dp}{ds} + F_z p + F_x &= 0 \\ \frac{dq}{ds} + F_z q + F_y &= 0. \end{aligned} \tag{A.3.13}$$

In summary: If a focal curve is embedded in an integral surface then along this curve the coordinates x, y, z and the quantities p and q satisfy the following five ordinary differential equations:

$$\begin{aligned} &\frac{dx}{ds} = F_p \qquad \frac{dy}{ds} = F_q \qquad &&\frac{dz}{ds} = pF_p + qF_q \\ &\frac{dp}{ds} = -(pF_z + F_x) &&\frac{dq}{ds} = -(qF_z + F_y). \end{aligned} \tag{A.3.14}$$

We can consider this system of differential equations (which are called characteristic equations) by itself, i.e., disregarding that we obtained it with a given integral surface in mind. Note first that $F(x, y, z, p, q)$ is constant along each solution to the system (A.3.14) since:

$$\begin{aligned} \frac{dF}{ds} =& F_p \frac{dp}{ds} + F_q \frac{dq}{ds} + F_z \frac{dz}{ds} + F_x \frac{dx}{ds} + F_y \frac{dy}{ds} = \\ &= - F_p(pF_z + F_x) - F_q(qF_z + F_y) + F_z(pF_p + qF_q) + F_x F_p + F_y F_q \\ &= 0. \end{aligned} \tag{A.3.15}$$

AD-A104 043 MASSACHUSETTS INST OF TECH CAMBRIDGE ARTIFICIAL INTE--ETC F/G 12/1
THE IMAGE IRRADIANCE EQUATION: ITS SOLUTION AND APPLICATION.(U)
JUN 81 A R BRUSS N00014-80-C-0505
UNCLASSIFIED AI-TR-623 NL

2 of 2
AI
A 104 043

END
DATE
FILMED
10-81
DTIC

Thus $F(x,y,z,p,q) = c$, where c is a constant, is a solution to (A.3.14). The system of characteristic equations defines a four-parameter family of curves. By imposing the additional constraint that the solutions to (A.3.14) also satisfy the FOPDE $F(x,y,z,p,q) = 0$ we obtain a three-parameter subset of the solutions. "Every solution of the characteristic differential equations which also satisfies the equation $F = 0$ will be called a *characteristic strip*; a space curve $x(s), y(s), z(s)$ bearing such a strip is called a characteristic curve" [COHI62b, p.79]. The fact that a one-parameter subset of the three-parameter family of curves sweeps out the integral surface has already been established. Thus the problem of finding a solution to a FOPDE is equivalent to integrating the system of five (or equivalently four if the equations are not written in parametric form) ordinary differential equations (A.3.14). Note that since the characteristic curves depend on the solution, their range of influence cannot be determined in advance.

In the next section we discuss the notion of a complete integral and then show how to choose the appropriate one-parameter subset of the solutions to the characteristic equations.

A.4. General Solution, Complete and Singular Integral

We showed in the previous section that each solution to a general FOPDE is swept out by a one-parameter family of curves. Thus we can write the equation of an integral surface as a function of the coordinates x and y and an arbitrary function of one variable. Such an equation is called the general solution to a FOPDE.

Suppose now for a moment that a solution $z = \Phi(x,y,a,b)$ to a FOPDE is known, where Φ depends on the two parameters a and b. Then $\Phi(x,y,a,b)$ is called a complete integral if Δ, defined by:

$$\Delta = \Phi_{xa}\Phi_{yb} - \Phi_{xb}\Phi_{ya} \tag{A.4.1}$$

is not equal to zero. This condition assures that Φ really depends on a and b, i.e., that there is no $\gamma = g(a,b)$ such that $\Phi(x,y,a,b) = \Phi(x,y,\gamma)$.

From the two-parameter family of surfaces defined by $\Phi(x,y,a,b)$, we can chose a one-parameter subset by introducing an arbitrary function w which relates a and b, e.g., by setting $b = w(a)$. Note that the family $\Phi(x,y,a,w(a))$ is a solution to the FOPDE. The envelope of the family $\Phi(x,y,a,w(a))$ is again a solution to the FOPDE since at each point it touches a member of the family $\Phi(x,y,a,w(a))$, i.e., a solution. We obtain the equation of this envelope by eliminating the parameter a from the two equations:

$$\begin{aligned} z &= \Phi(x,y,a,w(a)) \\ \Phi_a(x,y,a,w(a)) + \Phi_b(x,y,a,w(a))w'(a) &= 0. \end{aligned} \tag{A.4.2}$$

We assume throughout that all eliminations are possible and that during the course of this process only functions with continuous derivatives are obtained. Eliminating a

from (A.4.1) yields an expression involving an arbitrary function w of one variable which is a solution to the FOPDE and therefore the general solution. We demonstrate this fact analytically by differentiating the first equation of (A.4.2) with respect to x and y:

$$z_x = \Phi_x + (\Phi_a + \Phi_b w'(a))a_x \tag{A.4.3}$$
$$z_y = \Phi_y + (\Phi_a + \Phi_b w'(a))a_y.$$

Recall that $\Phi(x, y, a, w(a))$ is a solution to the FOPDE for any choice of the parameter a. Using (A.4.2) (i.e., $\Phi_a + \Phi_b w'(a) = 0$) in the previous equations establishes the fact that for all x and y the values of z, z_x and z_y are the same as those of Φ, Φ_x and Φ_y.

So if a complete integral of a given FOPDE is known, we can obtain the general solution by differentiation and by elimination of parameters. (This latter process can in practice be tedious or impossible but is often not necessary since every solution to the FOPDE is obtained by substituting all possible values for a.) In the next section we will show that with the help of the characteristic equations, a complete integral can found.

The general solution does not comprise all solutions to a FOPDE. The envelope of a complete integral, the so called *singular integral*, is a solution which cannot be obtained from the general solution. The equation of the singular integral, which does not contain any arbitrary elements, is found by eliminating the parameters a and b from the equations:

$$\begin{aligned} \Phi(x, y, a, b) &= z \\ \Phi_a(x, y, a, b) &= 0 \\ \Phi_b(x, y, a, b) &= 0. \end{aligned} \tag{A.4.4}$$

In fact, we do not have to know a complete integral of a FOPDE in order to find the singular solution. Note that for a complete integral Φ, $F(x, y, \Phi, \Phi_x, \Phi_y)$ vanishes identically for all choices of the parameters a and b. Differentiating the FOPDE with respect to a and b we obtain:

$$F_\Phi \Phi_a + F_p \Phi_{xa} + F_q \Phi_{ya} = 0 \tag{A.4.5}$$
$$F_\Phi \Phi_b + F_p \Phi_{xb} + F_q \Phi_{yb} = 0.$$

As Φ is a complete integral, $\Delta = \Phi_{xa}\Phi_{yb} - \Phi_{xb}\Phi_{ya}$ is not equal to zero. Furthermore Φ_a and Φ_b are zero (equations (A.4.4)) and therefore we may derive the equation of the singular integral by eliminating p and q from:

$$F_p = 0 \qquad F_q = 0 \qquad F = 0. \tag{A.4.6}$$

Note that equations (A.4.6) are derived from (A.4.5). (Remark: In this case we do not assume that $F_p^2 + F_q^2 \neq 0$ as we did to obtain the characteristic equations.)

If a FOPDE does not contain the function $z(x, y)$ explicitly, then no singular solution can exist as in this case the complete integral is of the form [COH62b, p.24]:

$$z = \Phi(x, y, a) + b \tag{A.4.7}$$

and the condition $\Phi_b = 0$ cannot be fulfilled.

A.5. A Method for Finding the Complete Integral

In the previous sections we showed two methods for finding the solutions to a given FOPDE. Here, we explain how to find the complete integral using the characteristic equations. Furthermore we describe how to determine a one-parameter subset from the four-parameter family of solutions to the characteristic equations.

First let us discuss a special form of a FOPDE called Pfaff's equation:

$$f(x,y,z)dx + g(x,y,z)dy + h(x,y,z)dz = 0. \tag{A.5.1}$$

In the case where $h \equiv 0$ and f and g depend only on x and y, this equation degenerates to an ordinary differential equation called an *exact* differential equation:

$$f(x,y)dx + g(x,y)dy = 0. \tag{A.5.2}$$

The equation is called *total* if f and g satisfy the integrability condition:

$$f_y(x,y) = g_x(x,y). \tag{A.5.3}$$

In the case of a total differential equation it is easy to find a solution to (A.5.2). On each simply connected region we can determine a function $H(x,y)$ such that:

$$\frac{\partial H}{\partial x} = f(x,y) \qquad \text{and} \qquad \frac{\partial H}{\partial y} = g(x,y). \tag{A.5.4}$$

Then:

$$dH = f(x,y)dx + g(x,y)dy \tag{A.5.5}$$

and the equation $dH = 0$ are both equivalent to (A.5.2). Thus $H(x,y) = c$, where c is a constant, is a solution to (A.5.2) and the function H can be found by integrating (A.5.4).

In the case where a FOPDE is exact but not total, an *integration factor* $\mu(x,y)$ can be always introduced such that the equation:

$$\mu f dx + \mu g dy = 0 \tag{A.5.6}$$

is total, i.e., such that $(\mu f)_y = (\mu g)_x$. Equivalently, $\mu(x,y)$ has to be a solution to the following FOPDE which we can solve using the method of characteristic curves:

$$\mu(f_y - g_x) + \mu_y f - \mu_x g = 0. \tag{A.5.7}$$

Equation (A.5.1) is also easy to solve if its left hand side is a total differential of a function $H(x,y,z)$ i.e., if:

$$f = \frac{\partial H}{\partial x} \qquad g = \frac{\partial H}{\partial y} \qquad h = \frac{\partial H}{\partial z}. \tag{A.5.8}$$

For the previous equations to hold, i.e., for the functions f, g and h to be the first order partial derivatives of a function H, it is necessary that the following integrability conditions be satisfied:

$$\frac{\partial f}{\partial y}=\frac{\partial g}{\partial x} \qquad \frac{\partial g}{\partial z}=\frac{\partial h}{\partial y} \qquad \frac{\partial h}{\partial x}=\frac{\partial f}{\partial z}. \tag{A.5.9}$$

In a simply connected region these integrability conditions assure the existence of a function $H(x,y,z)$ which satisfies (A.5.8) and can be calculated as:

$$H(x,y,z)=\int_{(x_0,y_0,z_0)}^{(x,y,z)}(f\,dx+g\,dy+h\,dz)+C \tag{A.5.10}$$

where (x_0,y_0,z_0) is a fixed point and C is a constant. Clearly $H(x,y,z)=c$, where c is a constant, is a solution to (A.5.1).

In (A.5.9) is not satisfied, it is again desirable to find an integration factor $\mu(x,y,z)$ such that the expression $\mu f dx+\mu g dy+\mu h dz$ is a total differential of a function. In contrast to the case of Pfaff's equation in two variables, it is not always possible to find such a factor. For such a μ to exist, it is necessary that the following equation holds:

$$f(g_z-h_y)+g(h_x-f_z)+h(f_y-g_x)=0. \tag{A.5.11}$$

It can also be shown (simple, but tedious) that in a simply connected region the previous equation is sufficient for (A.5.1) to possess a one-parameter family of solutions $H(x,y,z)=c$. We demonstrate now how to construct such a function $H(x,y,z)$.

First consider the *abbreviated* equation:

$$f(x,y,z)dx+g(x,y,z)dy=0. \tag{A.5.12}$$

This is a Pfaff's equation in the two variables x and y, with z as a parameter. Thus a solution to it can be always found (if necessary with the help of an integrating factor $\lambda(x,y,z)$):

$$\Phi(x,y)=u(x,y,z)=c \tag{A.5.13}$$

where c is a constant. Note that:

$$\lambda f=\frac{\partial u}{\partial x} \quad \text{and} \quad \lambda g=\frac{\partial u}{\partial y}. \tag{A.5.14}$$

Now we define a function S which depends upon the three variables x,y and z:

$$S(x,y,z)=\lambda h-\frac{\partial u}{\partial z}. \tag{A.5.15}$$

Using (A.5.13) we redefine S as a function of x, u and z, say:

$$T(x,u,z) = S(x,y,z). \tag{A.5.16}$$

Suppose that T is independent of x. Then we can find $H(x,y,z)$ by solving another Pfaff's equation in the variables u and z. In fact, it is easy to see that T is independent of x. One just has to prove that $\frac{\partial T}{\partial x} = 0$ which we do using equations (A.5.14) and (A.5.15):

$$\begin{aligned}\frac{\partial}{\partial x}(\lambda h - S) &= u_{xz} = \frac{\partial}{\partial z}(\lambda f)\\ \frac{\partial}{\partial y}(\lambda h - S) &= u_{yz} = \frac{\partial}{\partial z}(\lambda g)\\ \frac{\partial}{\partial x}(\lambda g) &= u_{xy} = \frac{\partial}{\partial y}(\lambda f).\end{aligned} \tag{A.5.17}$$

Equations (A.5.17) written out in full are:

$$S_x = h\lambda_x - f\lambda_z + \lambda(h_x - f_z) \tag{A.5.18}$$

$$-S_y = g\lambda_z - h\lambda_y + \lambda(g_z - h_y) \tag{A.5.19}$$

$$0 = f\lambda_y - g\lambda_x + \lambda(f_y - g_x). \tag{A.5.20}$$

Multiplying (A.5.18) by g, (A.5.19) by f, and (A.5.20) by h and then adding up the three equations using condition (A.5.11) yields:

$$gS_x - fS_y = 0. \tag{A.5.21}$$

Differentiating (A.5.16) with respect to x and y gives:

$$\begin{aligned}S_x &= T_x + T_u u_x\\ S_y &= T_u u_y.\end{aligned} \tag{A.5.22}$$

By combining (A.5.14), (A.5.21) and (A.5.22), we obtain:

$$\begin{aligned}0 = gS_x - fS_y &= gT_x + gT_u u_x - fT_u u_y =\\ &= gT_x + \lambda fgT_u - \lambda fgT_u = gT_x.\end{aligned} \tag{A.5.23}$$

Because $g \neq 0$ we may conclude that $T_x = 0$. Thus we rewrite (A.5.16) as:

$$S(x,y,z) = T(u,z). \tag{A.5.24}$$

Equation (A.5.1), after multiplication by λ and using the expressions (A.5.14) and (A.5.15) becomes:

$$\lambda(fdx + gdy + hdz) = u_x dx + u_y dy + (u_z + T)dz = 0 \tag{A.5.25}$$

or, equivalently:

$$du + T(u, z)dz = 0. \qquad \text{(A.5.26)}$$

This is again a Pfaff's equation in two variables, which can always be solved using an integrating factor. Its solution is $\psi(u, z) = c$ where c is a constant. Therefore the solution to (A.5.1) is:

$$H(x, y, z) = \psi(u(x, y, z), z) = c \qquad \text{(A.5.27)}$$

which is a one-parameter manifold.

Finally, we describe a method for finding a complete integral of a general FOPDE $F(x, y, z, p, q) = 0$. The basic idea is that the total differential of a solution $z(x, y)$ to the FOPDE:

$$dz = pdx + qdy \qquad \text{(A.5.28)}$$

can be interpreted as Pfaff's equation in the variables x, y and z. In order to do so, p and q must be expressed as functions of x, y and z. So assume that two functions f and g can be found such that:

$$p = f(x, y, z, a) \qquad \text{and} \qquad q = g(x, y, z, a) \qquad \text{(A.5.29)}$$

where a is an arbitrary constant. Then the FOPDE and (A.5.11) (with $h \equiv -1$):

$$fg_z - gf_z - f_y + g_x = 0 \qquad \text{(A.5.30)}$$

are satisfied. In this case the solution to (A.5.12) is a one-parameter manifold, but since a parameter a is already build into the equation, this solution contains two parameters and is the complete integral. Hence the problem of finding the appropriate functions f and g has to be solved.

Suppose that a function $G(x, y, z, p, q)$ exists such that the following equation can be solved for p and q (or equivalently for f and g):

$$\begin{aligned} F(x, y, z, p, q) &= 0 \qquad \text{(A.5.31)} \\ G(x, y, z, p, q) &= a. \end{aligned}$$

For this to be possible, the inequality $F_pG_q - F_qG_p \neq 0$ has to hold. We have to prove that f and g obtained in such a fashion satisfy (A.5.30) identically in the three variables x, y and z. Differentiating the equations (A.5.31) with respect to x, y and z gives:

$$\begin{aligned} F_x + p_xF_p + q_xF_q &= 0 \qquad & G_x + p_xG_p + q_xG_q &= 0 \\ F_y + p_yF_p + q_yF_q &= 0 \qquad & G_y + p_yG_p + q_yG_q &= 0 \qquad \text{(A.5.32)} \\ F_z + p_zF_p + q_zF_q &= 0 \qquad & G_z + p_zG_p + q_zG_q &= 0. \end{aligned}$$

After expressing p_z, q_z, p_y and q_x from the previous equations in terms of the derivatives of F and G and substituting these expressions into (A.5.30) we obtain a linear FOPDE

for the function G:

$$F_pG_x + F_qG_y + (pF_p + qF_q)G_z - (F_x + pF_z)G_p - (F_y + qF_z)G_q = 0. \quad \text{(A.5.33)}$$

We can solve this equation by the method of characteristic curves. The appropriate system of characteristic equations is the same as the one for $F(x, y, z, p, q) = 0$:

$$dx{:}dy{:}dz{:}dp{:}dq = F_p{:}F_q{:}(pF_p + qF_q){:} - (pF_z + F_x){:} - (qF_z + F_y). \quad \text{(A.5.34)}$$

Only one solution to these equations which is independent of F and contains at least one of the variables p and q is needed. Such an integral is the desired function G and will always exist since the solution to the characteristic equations comprises a four-parameter family:

$$v_i(x, y, z, p, q) = C_i \qquad i = 1, 2, 3, 4. \quad \text{(A.5.35)}$$

The v_i are independent and at least one of them must contain either p or q.

The method we have described is due to Lagrange and Charpit. It has the advantage over the method of characteristic curves discussed in section A.4 in that one needs only to find a single integral of (A.5.34) rather then a four-parameter family of curves.

A.6. The Initial Value Problem for Quasi-Linear FOPDE's

In this section we attack the problem of how to determine a *particular* solution of a FOPDE, once the general solution is known. First we consider a quasi-linear FOPDE:

$$a(x, y, z)p + b(x, y, z)q = c(x, y, z). \quad \text{(A.6.1)}$$

In particular we discuss the problem of how to find an integral surface $z(x, y)$ of (A.6.1) which contains a given curve C in space. In the literature, e.g., [COHI62b, p. 40], this problem is referred to as Cauchy's problem. Clearly the following questions have to be answered:

1) What conditions on C are necessary to make this problem is solvable?

2) When is such a solution unique?

Let C be given by three continuous differentiable functions of a parameter t: $x(t), y(t), z(t)$. Furthermore, we assume that the projection of C onto the x-y plane (referred to as C_0) does not contain double points and that $x_t^2 + y_t^2 \neq 0$. If C_0 contains double points, an integral surface with self-intersections is obtained, hence, z is not everywhere a single-valued function of x and y which implies that along the line of intersection, p and q

are discontinuous. Now to find a solution to the FOPDE containing C, we construct a characteristic curve through each point of C. The equations for the characteristic curves depend upon two parameters:

$$x = x(s,t) \qquad y = y(s,t) \qquad z = z(s,t). \tag{A.6.2}$$

Note that the functions x, y, z are still continuously differentiable. To get the equation of the integral surface, we must eliminate the parameters s and t from the previous equations, i.e., we must express s and t in terms of x and y. A sufficient condition for this to be possible is that the functional determinant Δ, which is specified by the following equation, does not vanish along the curve C:

$$\Delta = \frac{dx}{ds}\frac{dy}{dt} - \frac{dy}{ds}\frac{dx}{dt}. \tag{A.6.3}$$

Using the characteristic equations, we rewrite (A.6.3) as $\Delta = a\frac{dy}{dt} - b\frac{dx}{dt}$. Thus if $\Delta \neq 0$, we may express z as a function of x and y and be assured that C lies on the surface. This solution is also unique which is a consequence of the following lemma:

> **Lemma** [COHI62b, p.64]: Each characteristic curve which has one point in common with an integral surface lies completely on this surface.

This lemma follows from the uniqueness theorem for solutions to ordinary differential equations.

We can interpret the determinant Δ as the outer product of the two vectors:

$$\xi_1 = \begin{pmatrix} \frac{dx}{ds} \\ \frac{dy}{ds} \end{pmatrix} \quad \text{and} \quad \xi_2 = \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix} \tag{A.6.4}$$

which are the projections of the tangent and the characteristic direction onto the x-y plane, respectively. In the special case where Δ vanishes along C, these two directions coincide and we may deduce that C has one of the following three properties:

1) C is a characteristic curve.

2) C is the envelope of the characteristic curves (called an edge of regression).

3) C_0 is the envelope of the projections of the characteristic curves onto the x-y plane.

We discuss when case 1 occurs first. From $\Delta = 0$ we infer that:

$$\frac{1}{a}\frac{dx}{dt} = \frac{1}{b}\frac{dy}{dt}. \tag{A.6.5}$$

Using $x(t)$ and $y(t)$ (from the equation for C) in $z(x,y)$ the following equation has to hold along C:

$$\frac{dz}{dt} = z_x \frac{dx}{dt} + z_y \frac{dy}{dt} = ap + bq = c. \tag{A.6.6}$$

This means that C satisfies the characteristic equations and is therefore a characteristic curve. In this case, there exist infinitely many surfaces through C which satisfy the FOPDE. To see this, just choose another curve $\tilde{C}$ which has a point P in common with C. Now to construct the solutions through $\tilde{C}$, a characteristic curve is passed through every point of $\tilde{C}$, in particular through P. The characteristic curve through P is C, thus an integral surface through $\tilde{C}$ contains C. In this manner we can construct infinitely many integral surfaces which contain C. They all meet along the characteristic curve C which therefore can be viewed as a branch curve.

One assumption made throughout should be stressed again here: we require the solutions to a FOPDE to be continuous and continuously differentiable in some neighborhood of C. It might be possible to find a solution z through C, along which Δ vanishes, without C being a characteristic curve as occurs in cases 2 or 3 mentioned above. However, the derivatives of z are then not continuous along C. This fact is illustrated in the following example.

Example:

We wish to solve the following equation:

$$F(x,y,z,p,q) = 3(z-y)^2 p - q = 0. \tag{A.6.7}$$

The characteristic equations are:

$$\begin{aligned} \frac{dx}{ds} &= 3(z-y)^2 \\ \frac{dy}{ds} &= -1 \\ \frac{dz}{ds} &= 0. \end{aligned} \tag{A.6.8}$$

The solution to these equations, with the initial values x_0, y_0, z_0, is:

$$\begin{aligned} x &= (z_0 - y_0 + s)^3 + x_0 - (z_0 - y_0)^3 \\ y &= -s + y_0 \\ z &= z_0. \end{aligned} \tag{A.6.9}$$

Now we impose the constraint that the integral surface passes through C which is given by:

$$x = 0 \qquad y = t \qquad z = t. \tag{A.6.10}$$

Note that C is not a characteristic curve. Setting the initial values x_0, y_0, z_0 (i.e., x, y, z for $s = 0$) to:

$$x_0 = 0 \qquad y_0 = t \qquad z_0 = t, \tag{A.6.11}$$

(A.6.9) becomes:

$$\begin{aligned} x &= s^3 \\ y &= -s + t \\ z &= t. \end{aligned} \tag{A.6.12}$$

In this case the determinant Δ is:

$$\Delta = \frac{dx}{ds}\frac{dy}{dt} - \frac{dx}{dt}\frac{dy}{ds} = 3s^2. \tag{A.6.13}$$

Thus along the curve C (i.e., $s = 0$) $\Delta = 0$.

There is, however, a solution to the FOPDE which contains C:

$$z = x^{\frac{1}{3}} + y. \tag{A.6.14}$$

Note that $p = \frac{1}{3}x^{-\frac{2}{3}}$ does not exist along C (as $x = 0$ there).

In the case of a linear FOPDE we can make some further statements about the solution in the case where Δ vanishes along C. Here, the integral surfaces are cylindrical surfaces perpendicular to the x-y plane, i.e., the function defining an integral surface is independent of z. The linear FOPDE is:

$$a(x,y)p + b(x,y)q = c(x,y). \tag{A.6.15}$$

Recall that Δ is defined as:

$$\Delta = x_s y_t - x_t y_s. \tag{A.6.16}$$

Using the characteristic equations:

$$\begin{aligned} x_s &= a \\ y_s &= b \end{aligned} \tag{A.6.17}$$

the following equation is obtained for Δ_s:

$$\begin{aligned} \Delta_s &= a_s y_t + a y_{st} - b_s x_t - b x_{st} = \\ &= a_s y_t + a b_t - b_s x_t - a_t b. \end{aligned} \tag{A.6.18}$$

Note that if Δ_s (the first order partial derivative of Δ with respect to s) and Δ vanish along C, then Δ vanishes everywhere. The proof of this last assertion follows from the

existence and uniqueness theorem for ordinary differential equations. By differentiating a and b with respect to s and t and using relations (A.6.17), equations (A.6.19) are obtained:

$$\begin{aligned} a_s &= a_x a + a_y b \\ a_t &= a_x x_t + a_y y_t \\ b_s &= b_x a + b_y b \\ b_t &= b_x x_t + b_y y_t \end{aligned} \tag{A.6.19}$$

and so:

$$\Delta_s = (a_x + b_y)\Delta. \tag{A.6.20}$$

To express x as a function of y and z, say, $x = f(y,z)$, we have to assume that $y_s z_t - z_s y_t \neq 0$ along C.

Now, in order to prove that the integral surface z is a cylindrical surface, it suffices to show that $f_z = 0$. Differentiating x with respect to s and t we derive the following equations:

$$\begin{aligned} x_s &= f_y y_s + f_z z_s \\ x_t &= f_y y_t + f_z z_t. \end{aligned} \tag{A.6.21}$$

Then:

$$\Delta = f_z(z_s y_t - z_t y_s) \tag{A.6.22}$$

from which it follows that:

$$f_z = 0. \tag{A.6.23}$$

A.7. The Initial Value Problem for General FOPDE's

Here we pose the question: What are the constraints necessary to determine a solution to a general FOPDE uniquely? Clearly more information than in the quasi-linear case is needed as now the solutions to the characteristic equations form a three-parameter family of curves. So let C be a curve given by $x(t), y(t)$ and $z(t)$ such that neither C nor its projection onto the x-y plane have double points. Furthermore $p(t)$ and $q(t)$ along C have to be specified such that the condition:

$$\frac{dz}{dt} = p\frac{dx}{dt} + q\frac{dy}{dt} \tag{A.7.1}$$

holds and the FOPDE (A.3.1) (i.e., $F = 0$) is identically satisfied in t. Thus the functions $x(t), y(t), z(t), p(t)$ and $q(t)$ define an initial strip which we denote by C_1. From now on, the procedure is very similar to the one used in solving the initial value problem for a quasi-linear FOPDE. Through every element of C_1 a characteristic strip

is constructed which can be written as $x(s,t), y(s,t), z(s,t), p(s,t)$ and $q(s,t)$. To be able to express the parameters s and t in terms of x and y:

$$\Delta = x_s y_t - x_t y_s = F_p y_t - F_q x_t \tag{A.7.2}$$

cannot vanish identically along the initial strip. Assuming this holds, z, p and q can be expressed in terms of x and y. It then remains to check whether p and q written in such a fashion are the first order partial derivatives of the integral surface $z(x,y)$. This involves showing that the quantities U and V:

$$\begin{aligned} U &= z_t - p x_t - q y_t \\ V &= z_s - p x_s - q y_s \end{aligned} \tag{A.7.3}$$

vanish identically. Since we assumed that $\Delta \neq 0$ we deduce from the previous equations and:

$$\begin{aligned} 0 &= z_t - z_x x_t - z_y y_t \\ 0 &= z_s - z_x x_s - z_y y_s \end{aligned} \tag{A.7.4}$$

that $z_x = p$ and $z_y = q$. Recall now the characteristic equations:

$$\frac{dx}{ds} = F_p \qquad \frac{dy}{ds} = F_q \qquad \frac{dz}{ds} = pF_p + qF_q. \tag{A.7.5}$$

Using the first two of these equations in the last one, we obtain:

$$\frac{dz}{ds} = p\frac{dx}{ds} + q\frac{dy}{ds} \tag{A.7.6}$$

which implies that V vanishes identically. Now to prove that also U vanishes identically, note that:

$$\frac{\partial U}{\partial s} = z_{st} - p_s x_t - p x_{st} - q_s y_t - q y_{st} \tag{A.7.7}$$

$$\frac{\partial V}{\partial s} = z_{st} - p_t x_s - p x_{st} - q_t y_s - q y_{st}. \tag{A.7.8}$$

Subtracting (A.7.7) from (A.7.8) yields:

$$\frac{\partial U}{\partial s} - \frac{\partial V}{\partial t} = -p_s x_t - p_t x_s + q_s y_t - q_t y_s. \tag{A.7.9}$$

Taking into account the characteristic equations and the fact that $V \equiv 0$ implies $\frac{\partial V}{\partial t} = 0$, we rewrite the previous equation as:

$$\frac{\partial U}{\partial s} = p_t F_p + q_t F_q + x_t F_x + y_t F_y + (p x_t + q y_t) F_z. \tag{A.7.10}$$

However, the FOPDE $F = 0$ holds identically in s and t. Differentiating F with respect to t gives:

$$F_x x_t + F_y y_t + F_z z_t + F_p p_t + F_q q_t = 0. \tag{A.7.11}$$

Using the previous equation in (A.7.10) we obtain:

$$\frac{\partial U}{\partial s} = -F_z U. \tag{A.7.12}$$

For any fixed t, (A.7.12) is an ordinary differential equation for U as a function of s with the solution:

$$U(s) = U(0)e^{\int_0^s -F_z U}. \tag{A.7.13}$$

Since by assumption, $U(0)$ is zero, U vanishes everywhere.

To summarize the previous results: given a curve $x(t), y(t)$ and $z(t)$ along which $p(t)$ and $q(t)$ are known such that:

$$\begin{aligned} \frac{dz}{dt} &= p\frac{dx}{dt} + q\frac{dy}{dt} \\ F(x(t), y(t), z(t), p(t), q(t)) &= 0 \\ \Delta = F_p y_t - F_q x_t &\neq 0, \end{aligned} \tag{A.7.14}$$

there exists a unique integral surface through the initial strip. We obtain a unique surface because the solution to the characteristic equations is uniquely determined by their initial values.

The exceptional case where $\Delta = 0$ along C_1 is analogous to the one discussed in the previous section: there are infinitely many integral surfaces through C_1 if and only if it is a characteristic strip. Again we can view any characteristic curve as a branch element since, on either side of it, there can be another member of the family of solutions to a FOPDE, while along such a curve the first derivatives are continuous. Note that higher order derivatives along the curve may be discontinuous. If C_1 is only a focal strip along which $\Delta = 0$, then it might be possible to find an integral surface z containing it. As in the quasi-linear case, this surface does not have continuous derivatives.

Finally, suppose C degenerates to a point P with coordinates (x_0, y_0, z_0). Then the strip condition is identically satisfied for all p_0 and q_0 which satisfy the FOPDE, i.e., for all p_0 and q_0 which determine the feasible tangent planes at P. So p_0 and q_0 can be written as functions of a parameter t. If the quantities $x_0, y_0, z_0, p_0(t)$ and $q_0(t)$ are used as initial values when solving the characteristic equations, a unique integral surface which has a conical singularity at P is obtained. It is called the *integral conoid* of the partial differential equation at P.

Appendix II

Transformation

In this appendix we show that any solution to an image irradiance equation of the form (B.1) can be obtained from the solutions to an image irradiance equation of the form (B.5). Let

$$f(Ap^2 + 2Bpq + Cq^2 + 2Dp + 2Eq) = E(x, y) \tag{B.1}$$

be an image irradiance equation where f is a bijection and A, B, C, D and E are real constants such that $\delta > 0$ and $\Delta S < 0$, where δ, Δ and S are defined by:

$$\begin{aligned} \delta &= AC - B^2 \\ \Delta &= \begin{vmatrix} A & B & D \\ B & C & E \\ D & E & 0 \end{vmatrix} \\ S &= A + C. \end{aligned} \tag{B.2}$$

As f is a bijection, equation (B.1) and the following transformed equation have the same solutions:

$$Ap^2 + 2Bpq + Cq^2 + 2Dp + 2Eq = f^{-1}(E(x, y)). \tag{B.3}$$

Equivalently, we can write the previous equation as:

$$Ap^2 + 2Bpq + Cq^2 + 2Dp + 2Eq = \tilde{E}(x, y). \tag{B.4}$$

Now let $z = z(x, y)$ be a solution to:

$$p^2 + q^2 = \tilde{E}(\tilde{x}, \tilde{y}) + c \tag{B.5}$$

where $\tilde{x}$, $\tilde{y}$ and c are defined by:

$$\begin{aligned} \tilde{x} &= \frac{B}{\sqrt{C}}x + \frac{\sqrt{AC - B^2}}{\sqrt{C}}y \\ \tilde{y} &= \sqrt{C}x \\ c &= A\alpha^2 + 2B\alpha\beta + C\beta^2 \end{aligned} \tag{B.6}$$

and α and β are defined by:

$$\begin{aligned} \alpha &= \frac{CD - BE}{AC - B^2} \\ \beta &= \frac{AE - BD}{AC - B^2}. \end{aligned} \tag{B.7}$$

Then $\tilde{z}(\tilde{x}, \tilde{y}) = z(x(\tilde{x}, \tilde{y}), y(\tilde{x}, \tilde{y})) - \alpha\tilde{x} - \beta\tilde{y}$ is a solution to (B.3).

Proof: The proof proceeds by diagonalization of a quadratic form. First we express x and y as functions of $\tilde{x}$ and $\tilde{y}$:

$$\begin{aligned} x &= \frac{\tilde{y}}{\sqrt{C}} \\ y &= \frac{C\tilde{x} - B\tilde{y}}{\sqrt{C(AC - B^2)}}. \end{aligned} \tag{B.8}$$

The first order partial derivatives of $\tilde{z}$ are abbreviated by $\tilde{p}$ and $\tilde{q}$:

$$\begin{aligned} \tilde{p} &= \frac{\partial \tilde{z}}{\partial \tilde{x}} \\ \tilde{q} &= \frac{\partial \tilde{z}}{\partial \tilde{y}}. \end{aligned} \tag{B.9}$$

In the following equations we express $\tilde{p}$ and $\tilde{q}$ in terms of p and q and vice versa:

$$\begin{aligned} \tilde{p} &= q\frac{\sqrt{C}}{\sqrt{AC - B^2}} - \alpha \\ \tilde{q} &= p\frac{1}{\sqrt{C}} - q\frac{B}{\sqrt{C(AC - B^2)}} - \beta \\ p &= \frac{B(\tilde{p} + \alpha) + C(\tilde{q} + \beta)}{\sqrt{C}} \\ q &= \frac{\sqrt{AC - B^2}}{\sqrt{C}}(\tilde{p} + \alpha). \end{aligned} \tag{B.10}$$

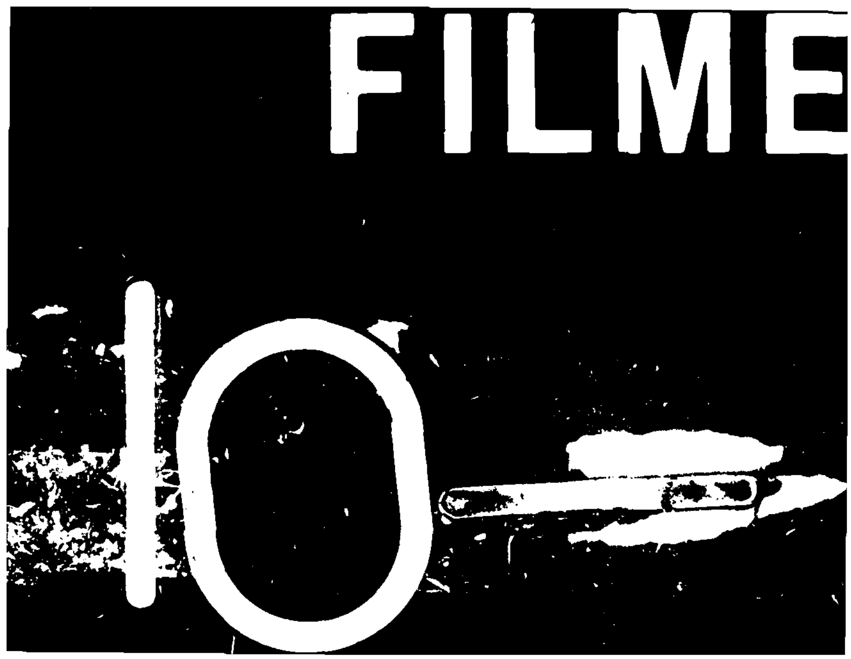
Thus:

$$\begin{aligned}
p^2 + q^2 = & \frac{1}{C}[B^2(\tilde{p}+\alpha)^2 + 2BC(\tilde{p}+\alpha)(\tilde{q}+\beta) + \\
& C^2(\tilde{q}+\beta)^2 + (AC - B^2)(\tilde{p}+\alpha)^2] \\
= \quad & A\tilde{p}^2 + 2A\alpha\tilde{p} + A\alpha^2 + 2B\tilde{p}\tilde{q} + 2B\alpha\tilde{q} + 2B\beta\tilde{p} + 2B\alpha\beta + \\
& C\tilde{q}^2 + 2C\beta\tilde{q} + C\beta^2 \\
= \quad & \tilde{E}(\tilde{x},\tilde{y}) + c.
\end{aligned} \tag{B.11}$$

From this last equation we deduce that:

$$A\tilde{p}^2 + 2B\tilde{p}\tilde{q} + C\tilde{q}^2 + 2D\tilde{p} + 2E\tilde{q} = E(\tilde{x},\tilde{y}) \tag{B.12}$$

which is the same as (B.4). ∎


FILME
10